General Internal Medicine
Boston University School of Medicine
2011 Publications – A-L

1. Agha G, **Murabito JM**, Lynch JW, Abrahamowicz M, Harper SB, Loucks EB. Am J Cardiol. Relation of socioeconomic position with ankle-brachial index. 2011 Dec;108(11):1651-7. PMID: 21907950.

2. **Alford DP**, **LaBelle CT**, Kretsch N, Bergeron A, Winter M, Botticelli M, **Samet JH**. Collaborative care of opioid-addicted patients in primary care using buprenorphine. Arch Intern Med. 2011;171(5):425-31.

3. **Battaglia TA**, Burhansstipanov L, Murrell SS, Dwyer AJ, Caron AS; Prevention and Early Detection Workgroup, National Patient Navigation Leadership Summit. Assessing the impact of patient navigation: Prevention and early detection metrics. Cancer. 2011 Aug;117(15 Suppl):3553-64. doi: 10.1002/cncr.26267. PMID: 21780090.

4. Berkowitz RE, Jones RN, Rieder R, Bryan M, Schreiber R, Verney S, **Paasche-Orlow MK**. Improving disposition outcomes for patients in a geriatric skilled nursing facility. J Am Geriatr Soc. 2011 Jun;59(6):1130-6. doi: 10.1111/j.1532-5415.2011.03417.x. PMID: 21649622.

5. **Berz JP**, Singer MR, Guo X, Daniels SR, Moore LL. Use of a DASH food group score to predict excess weight gain in adolescent girls in the National Growth and Health Study. Arch Pediatr Adolesc Med. 2011 Jun;165(6):540-6. PMID: 21646587.

6. Blood EA, **Cheng DM**. The use of mixed models for the analysis of mediated data with time-dependent predictors. J Environ Public Health. 2011;2011:435078. PMID: 21647351.

7. Blumhofer RD, Shah N, Grodin MA, **Crosby SS**. Clinical issues in caring for former chattel slaves. J Immigr Minor Health. Apr 2011;13(2):323-32. DOI:10.1007/s10903-008-9217-4.

8. **Borzecki AM**, Cevasco M, Chen Q, Shin M, Itani KM, Rosen AK. How valid is the AHRQ Patient Safety Indicator "postoperative physiologic and metabolic derangement"? J Am Coll Surg. 2011 Jun;212(6):968-976.e1-2. PMID: 21489834.

9. **Borzecki AM**, Kaafarani H, Cevasco M, Hickson K, Macdonald S, Shin M, Itani KM, Rosen AK. How valid is the AHRQ Patient Safety Indicator "postoperative hemorrhage or hematoma"? J Am Coll Surg. 2011 Jun;212(6):946-953.e1-2. PMID: 21474344.

10. **Borzecki AM**, Kaafarani HM, Utter GH, Romano PS, Shin MH, Chen Q, Itani KM, Rosen AK. How valid is the AHRQ Patient Safety Indicator "postoperative respiratory failure"? J Am Coll Surg. 2011 Jun;212(6):935-45. PMID: 21474343.

11. Cabral KP, Ansell J, **Hylek EM**. Future directions of stroke prevention in atrial fibrillation: The potential impact of novel anticoagulants and stroke risk stratification. J Thromb Haemost. 2011 Mar;9(3):441-9. PMID: 21176102. doi: 10.1111/j.1538-7836.2010.04179.x.

12. Carey K, Stefos T, Shibei Zhao, **Borzecki AM**, Rosen AK. Excess costs attributable to postoperative complications. Med Care Res Rev. 2011 Aug;68(4):490-503. PMID: 21536599.

13. Cevasco M, **Borzecki AM**, Chen Q, Zrelak PA, Shin M, Romano PS, Itani KM, Rosen AK. Positive predictive value of the AHRQ Patient Safety Indicator "Postoperative Sepsis": Implications for practice and policy. J Am Coll Surg. 2011 Jun;212(6):954-61. PMID: 21474345.

14. Cevasco M, **Borzecki AM**, McClusky DA 3rd, Chen Q, Shin MH, Itani KM, Rosen AK. Positive predictive value of the AHRQ Patient Safety Indicator "postoperative wound dehiscence". J Am Coll Surg. 2011 Jun;212(6):962-7. PMID: 21489829.

15. Cevasco M, **Borzecki AM**, O'Brien WJ, Chen Q, Shin MH, Itani KM, Rosen AK. Validity of the AHRQ Patient Safety Indicator "central venous catheter-related bloodstream infections". J Am Coll Surg. 2011 Jun;212(6):984-90. PMID: 21489833.

16. Chavez PR, Nelson DE, **Naimi TS**, Brewer RD. Impact of a new gender-specific definition for binge drinking on prevalence estimates for women. Am J Prev Med. 2011 Apr;40(4):468-71. PMID: 21406282. PMCID: PMC3090660.

17. Chaudhry AA, Berkenblit G, **Gifford AL**, Cofrancesco J Jr, Sosman J, Sullivan LE. Update in HIV medicine for the generalist. J Gen Intern Med. 2011 May;26(5):538-42. PMID: 21140297. PMCID: PMC3077485.

18. Chen Q, Rosen AK, Cevasco M, Shin M, Itani KM, **Borzecki AM**. Detecting patient safety indicators: How valid is "foreign body left during procedure" in the Veterans Health Administration? J Am Coll Surg. 2011 Jun;212(6):977-83. PMID: 21489830.

19. Coviello AD, Zhuang WV, Lunetta KL, Bhasin S, Ulloor J, Zhang A, Karasik D, Kiel DP, Vasan RS, **Murabito JM**. Circulating testosterone and SHBG concentrations are heritable in women: The Framingham Heart Study. 2011 Sep;96(9):E1491-5. PMID: 21752884.

20. **D'Amore MM**, **Cheng DM**, **Kressin NR**, Jones J, **Samet JH**, Winter M, **Kim TW**, **Saitz R**. Oral health of substance-dependent individuals: Impact of specific substances. J Subst Abuse Treat. 2011 Sep;41(2):179-85. PMID: 21474269.

21. De Caterina R, **Hylek EM.** Stroke prevention in atrial fibrillation: Current status and near-future directions. Am J Med. 2011;124:793-99. PMID: 21745652.

22. Dentali F, Gianni M, Squizzato A, Ageno W, Castiglioni L, Maroni L, **Hylek EM**, Grandi AM, Cazzani E, Venco A, Guasti L. Use of statins and recurrence of atrial fibrillation after catheter ablation or electrical cardioversion. A systematic review and meta-analysis. Thromb Haemost. 2011 Aug;106(2):363-70. PMID: 21614408.

23. Ding YY, Kader B, Christiansen CL, **Berlowitz DR**. Patient factors associated with transfusion practices in Veterans Affairs intensive care units: Implications for further research. J Crit Care. 2011 Aug;26(4):431.e1-9. PMID: 21376514.

24. Estes NA 3rd, Sacco RL, Al-Khatib SM, Ellinor PT, Bezanson J, Alonso A, Antzelevitch C, Brockman RG, Chen PS, Chugh SS, Curtis AB, Dimarco JP, Ellenbogen KA, Epstein AE, Ezekowitz MD, Fayad P, Gage BF, Go AS, Hlatky MA, **Hylek EM**, Jerosch-Herold M, Konstam MA, Lee R, Packer DL, Po SS, Prystowsky EN, Redline S, Rosenberg Y, Van Wagoner DR, Wood KA, Yue L, Benjamin EJ. American Heart Association atrial fibrillation research summit: a conference report from the american heart association. Circulation. 2011 Jul;124(3):363-72. PMID: 21709057. PMCID: PMC3144149.

25. **Farzanfar R**, Locke SE, Heeren TC, Stevens A, Vachon L, Nguyen MKT, **Friedman RH**. Workplace telecommunications technology to identify mental health disorders and facilitate self-help or professional referrals. Am J of Health Promotions. 2011 Jan-Feb;25(3)207-16.

26. Fiore LD, Brophy M, Ferguson RE, D'Avolio L, **Hermos JA**, Lew RA, Doros G, Conrad CH, O'Neil JA Jr, Sabin TP, Kaufman J, Swartz SL, Lawler E, Liang MH, Gaziano JM, Lavori PW. A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. Clin Trials. 2011;8(2):183-95. PMID: 21478329. doi: 10.1177/1740774510395635.

27. **Fleurant M**, Kell R, Love J, Jenter C, Volk LA, Zhang F, Bates DW, Simon SR. Massachusetts e-health project increased physicians' ability to use registries, and signals progress toward better care. Health Aff (Millwood). 2011 Jul;30(7):1256-64. PMID: 21734198.

28. **Freund KM**. Patient navigation: The promise to reduce health disparities. J Gen Intern Med. 2011 Feb;26(2):110-2. PMID: 21161422.

29. Garg S, Hoenig M, Edwards EM, Bliss C, Heeren T, Tumilty S, **Walley AY**, Koziel MJ, Skolnik PR, Horsburgh CR, Cotton D. Incidence and predictors of acute kidney injury in an urban cohort of subjects with HIV and hepatitis C virus coinfection. AIDS Patient Care STDS. 2011 Mar;25(3):135-41. PMID: 21309706.

30. Gordon AJ, Kunins HV, Rastegar DA, Tetrault JM, **Walley AY**. Update in addiction medicine for the generalist. J Gen Intern Med. 2011 Jan;26(1):77-82. PMID: 20697971. doi: 10.1007/s11606-010-1461-3. PMID: 20697971. PMCID: PMC3024103.

31. Granger CB, Alexander JH, McMurray JJ, Lopes RD, **Hylek EM**, Hanna M, Al-Khalidi HR, Ansell J, Atar D, Avezum A, Bahit MC, Diaz R, Easton JD, Ezekowitz JA, Flaker G, Garcia D, Geraldes M, Gersh BJ, Golitsyn S, Goto S, Hermosillo AG, Hohnloser SH, Horowitz J, Mohan P, Jansky P, Lewis BS, Lopez-Sendon JL, Pais P, Parkhomenko A, Verheugt FW, Zhu J, Wallentin L; ARISTOTLE Committees and Investigators. Apixaban versus warfarin in patients with atrial fibrillation. N Engl J Med. 2011 Sep 15;365(11):981-92. PMID: 21870978.

32. Groessl EJ, Weingart KR, **Gifford AL**, Asch SM, Ho SB. Development of the hepatitis C self-management program. Patient Educ Couns. 2011 May;83(2):252-5. PMID: 20638216.

33. Groessl EJ, Weingart KR, Stepnowsky CJ, **Gifford AL**, Asch SM, Ho SB. The hepatitis C self-management programme: A randomized controlled trial. J Viral Hepat. 2011 May;18(5):358-68. PMID: 20529203. doi: 10.1111/j.1365-2893.2010.01328.x.

34. Hanlon JT, Wang X, Castle NG, Stone RA, Handler SM, Semla TP, Pugh MJ, **Berlowitz DR**, Dysken MW. Potential underuse, overuse, and inappropriate use of antidepressants in older veteran nursing home residents. J Am Geriatr Soc. 2011 Aug;59(8):1412-20. doi: 10.1111/j.1532-5415.2011.03522.x. PMID: 21824120.

35. Haring R, Travison TG, Bhasin S, Vasan RS, Wallaschofski H, Davda MN, Coviello A, **Murabito JM**. Relation between sex hormone concentrations, Peripheral Arterial Disease, and change in Ankle-Brachial Index: Findings from the Framingham Heart Study. J Clin Endocrinol Metab. 2011 Dec;96(12):3724-32. PMID: 21824120.

36. Houlihan BV, Jette A, **Paasche-Orlow M**, Wierbicky J, Ducharme S, Zazula J, Cuevas P, **Friedman RH**, Williams S. A telerehabilitation intervention for persons with spinal cord dysfunction. Am J Phys Med Rehabil. 2011 Sep;90(9):756-64. PMID: 21389846. doi: 10.1097/PHM.0b013e31820b140f.

37. Ireland J, **Cheng DM**, **Samet JH**, **Bridden C**, Quinn E, **Saitz R**. Operating characteristics of carbohydrate-deficient transferrin (CDT) for identifying unhealthy alcohol use in adults with HIV infection. AIDS Care. 2011 Nov;23(11):1483-91. PMID: 21732900.

38. Jean-Pierre P, Fiscella K, **Freund KM**, Clark J, Darnell J, Holden A, Post D, Patierno SR, Winters PC; Patient Navigation Research Program Group. Structural and reliability analysis of a patient satisfaction with cancer-related care measure: A multisite patient navigation research program study. Cancer. 2011 Feb;117(4):854-61. PMID: 20922802. doi: 10.1002/cncr.25501.

39. Kaafarani H, **Borzecki, AM**, Hickson K, MacDonald S, Itani K, Shin M, Zhao S, Rosen AK. Validity of selected patient safety indicators: Opportunities and concerns. J Am Coll Surg. 2011 Jun;212(6):924-34. PMID 20869268.

40. Kabali C, **Cheng DM**, Brooks D, **Bridden C**, Horsburgh R Jr, **Samet JH**. Recent cigarette smoking and HIV disease progression: No evidence of an association. AIDS Care. 2011 Aug;23(8):947-56. PMID: 21400309.

41. Keating NL, O'Malley AJ, **Murabito JM**, Smith KP, Christakis NA. Minimal social network effects evident in cancer screening behavior. Cancer. 2011 Jul;117(13):3045-52. PMID: 21264828. PMCID: PMC3119780. doi: 10.1002/cncr.25849.

42. **Kim TW**, **Saitz R**, **Cheng DM**, Winter MR, **Witas J**, **Samet JH**. Initiation and engagement in chronic disease management care for substance dependence. Drug Alcohol Depend. 2011 May;115(1-2):80-6. PMID: 21168981.

43. Kraus ML, **Alford DP**, Kotz MM, Levounis P, Mandell TW, Meyer M, Salsitz EA, Wetterau N, Wyatt SA. Statement of the American Society of Addiction Medicine Consensus Panel on the use of buprenorphine in office-based treatment of opioid addiction. J Addict Med. 2011 Dec; 5(4):254-63. PMID: 22042215.

44. **Kronman AC**, **Paasche-Orlow M**, **Orlander JD**. Factors associated with disclosure of medical errors by housestaff. BMJ Qual Saf. 2011. doi:10.1136/bmjqs-2011-000084.

45. Lam CS, Cheng S, Choong K, Larson MG, **Murabito JM**, Newton-Cheh C, Bhasin S, McCabe EL, Miller KK, Redfield MM, Vasan RS, Coviello AD, Wang TJ. Influence of sex and hormone status on circulating natriuretic peptides. J Am Coll Cardiol. 2011 Aug;58(6):618-26. PMID: 21798425.

46. Langer EM, **Gifford AL**, Chan K. Comparative logic modeling for policy analysis: The case of HIV testing policy change at the Department of Veterans Affairs. Health Serv Res. 2011 Oct;46(5):1628-45. PMID: 21689094. doi: 10.1111/j.1475-6773.2011.01283.x.

47. **Lasser KE**, **Kim TW**, **Alford DP**, Cabral H, **Saitz R**, **Samet JH**. Is unhealthy substance use associated with failure to receive cancer screening and influenza vaccination? A retrospective cross-sectional study. BMJ Open. (online only) 2011:1-8.

48. **Lasser KE**, Murillo J, Lisboa S, Casimir N, Valley-Shah L, Fletcher RH, Emmons KE, Ayanian JZ. Colorectal cancer screening among ethnically diverse, low-income patients: A randomized controlled trial. Arch Intern Med. 2011 May;171(10):906-12. PMID: 21606094.

49. **Liebschutz JM**, **Alford DP**. Safe opioid prescribing: A long way to go. J Gen Intern Med. 2011;26(9):951-2. PMID: 21735346.

50. **Linsky A**, **Hermos JA**, Lawler EV, Rudolph, JL. Proton pump inhibitor discontinuation in long-term care. J Amer Geriatr Soc. 2011;59:1658-64. doi: 10.1111/j.1532-5415.2011.03545.x. PMID: 21883102.

51. **Linsky A**, Nyambose J, **Battaglia TA**. Lifestyle behaviors in Massachusetts cancer survivors. J Cancer Surviv. 2011 Mar;5(1):27-34. PMID: 21132395.

52. **Linsky A**, McIntosh N, Cabral H, Kazis L. Patient-provider language concordance and colorectal cancer screening. J Gen Intern Med. 2011 Feb;26(2):142-7. PMID: 20857340. PMCID: PMC3019323.

53. Lopes RD, Becker RC, Alexander JH, Armstrong PW, Califf RM, Chan MY, Crowther M, Granger CB, Harrington RA, **Hylek EM**, James SK, Jolicoeur EM, Mahaffey KW, Newby LK, Peterson ED, Pieper KS, Van de Werf F, Wallentin L, White HD, Carvalho AC, Giraldez RR, Guimaraes HP, Nader HB, Kalil RA, Bizzachi JM, Lopes AC, Garcia DA. Highlights from the III International Symposium of Thrombosis and Anticoagulation (ISTA), October 14-16, 2010, São Paulo, Brazil. J Thromb Thrombolysis. 2011 Aug;32(2):242-66. PMID: 21547405.

54. Lopes RD, Horowitz JD, Garcia DA, Crowther MA, **Hylek EM**. Warfarin and acetaminophen interaction: A summary of the evidence and biological plausibility. Blood. 2011 Dec;118(24):6269-73. PMID: 21911832.

55. **Lunze K**, Feldman L. Under pressure for a diagnosis. Clin Pediatr (Phila). 2011 Feb;50(2):166-8. PMID: 20118103. doi: 10.1177/0009922809350220.

56. **Lunze K**, Lunze FI. Addressing the burden of post-conflict surgical disease: Strategies from the North Caucasus. Glob Public Health. 2011 Sep;6(6):669-77. PMID: 21432701.

Relation of Socioeconomic Position With Ankle–Brachial Index

Golareh Agha, MSc[a,*], Joanne M. Murabito, MD[b,c], John W. Lynch, PhD[d,e], Michal Abrahamowicz, PhD[f], Sam B. Harper, PhD[f], and Eric B. Loucks, PhD[a]

Potential upstream determinants of coronary heart disease (CHD) include life-course socioeconomic position (e.g., childhood socioeconomic circumstances, own education and occupation); however, several plausible biological mechanisms by which socioeconomic position (SEP) may influence CHD are poorly understood. Several CHD risk factors appear to be more strongly associated with SEP in women than in men; little is known as to whether any CHD risk factors may be more strongly associated with SEP in men. Objectives were to evaluate whether cumulative life-course SEP is associated with a measurement of subclinical atherosclerosis, the ankle–brachial index (ABI), in men and women. This study was a prospective analysis of 1,454 participants from the Framingham Heart Study Offspring Cohort (mean age 57 years, 53.8% women). Cumulative SEP was calculated by summing tertile scores for father's education, own education, and own occupation. ABI was dichotomized as low (≤1.1) and normal (>1.1 to 1.4). After adjustment for age and CHD risk factors cumulative life-course SEP was associated with low ABI in men (odds ratio [OR] 2.04, 95% confidence interval [CI] 1.22 to 3.42, for low vs high cumulative SEP score) but not in women (OR 0.86, 95% CI 0.56 to 1.33). Associations with low ABI in men were substantially driven by their own education (OR 4.13, 95% CI 1.86 to 9.16, for lower vs higher than high school education). In conclusion, cumulative life-course SEP was associated with low ABI in men but not in women. **(Am J Cardiol 2011;108:1651–1657)**

Very little is known about whether life-course socioeconomic position (SEP) is associated with the ankle–brachial index (ABI), a subclinical measurement of atherosclerosis. ABI is the ratio of systolic blood pressure in the ankle to systolic blood pressure in the arm. It is well recognized that ABI is a measurement of generalized atherosclerosis and that an ABI <0.9 indicates peripheral arterial disease. In the Atherosclerosis Risk In Communities (ARIC)[1] study and in a systematic review including longitudinal studies from the United States and Europe,[2] low ABI was associated with increased incidence of total mortality, cardiovascular mortality, myocardial infarction, and stroke. The increased relative risks were shown to be independent of baseline cardiovascular disease and risk factors, suggesting that the ABI might have an independent role in predicting cardiovascular events. Consequently, the primary objective of this study was to investigate whether cumulative life-course SEP is associated with ABI in the Framingham Offspring Study cohort independent of classic coronary heart disease (CHD) risk factors.

Methods

The Framingham Heart Study is a community-based observational cohort study initiated in 1948 to investigate risk factors for CHD. The present investigation was based on participants in the Framingham Offspring Study, which began in 1971 with recruitment of 5,124 United States men and women who were offspring (or spouses of offspring) of the original cohort of the Framingham Heart Study. The design and selection criteria of the Framingham Offspring Study have been described elsewhere.[3] Participants were examined every 4 to 8 years, undergoing medical history, physical examination, anthropometry, and laboratory assessment of CHD risk factors at each examination, as previously described.[3] Framingham Study participants signed informed consent and the Framingham Study is reviewed annually by the Boston University Medical Center institutional review board.

Childhood SEP was measured by father's educational attainment in primary analyses and father's occupation in secondary analyses. Father's education and occupation were obtained directly from the participants' fathers who were enrolled in the Framingham Heart Study original cohort from 1948 through 1950 (mean age 44 years, range 28 to 62). Father's education was categorized as 3 levels: lower than high school, completed high school, and higher than high school. Father's occupation was categorized as 3 lev-

[a]Department of Epidemiology, Brown University, Providence, Rhode Island; [b]Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts; [c]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts; [d]School of Health Sciences, University of South Australia, Adelaide, Australia; [e]Department of Social Medicine, University of Bristol, Bristol, United Kingdom; [f]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada. Manuscript received March 20, 2011; revised manuscript received and accepted July 7, 2011.

The research was supported by Operating Grant MOP81239 from the Canadian Institutes of Health and Grant HOA80072 from the Canadian Institutes of Health Interdisciplinary Capacity Enhancement, Ottawa, Ontario, Canada. The Framingham Heart Study is supported by Contract N01-HC-25195 from the National Institutes of Health/National Heart, Lung, and Blood Institute, Bethesda, Maryland.

*Corresponding author: Tel: 401-261-3365.

E-mail address: Golareh_agha@brown.edu (G. Agha).

0002-9149/11/$ – see front matter
doi:10.1016/j.amjcard.2011.07.030

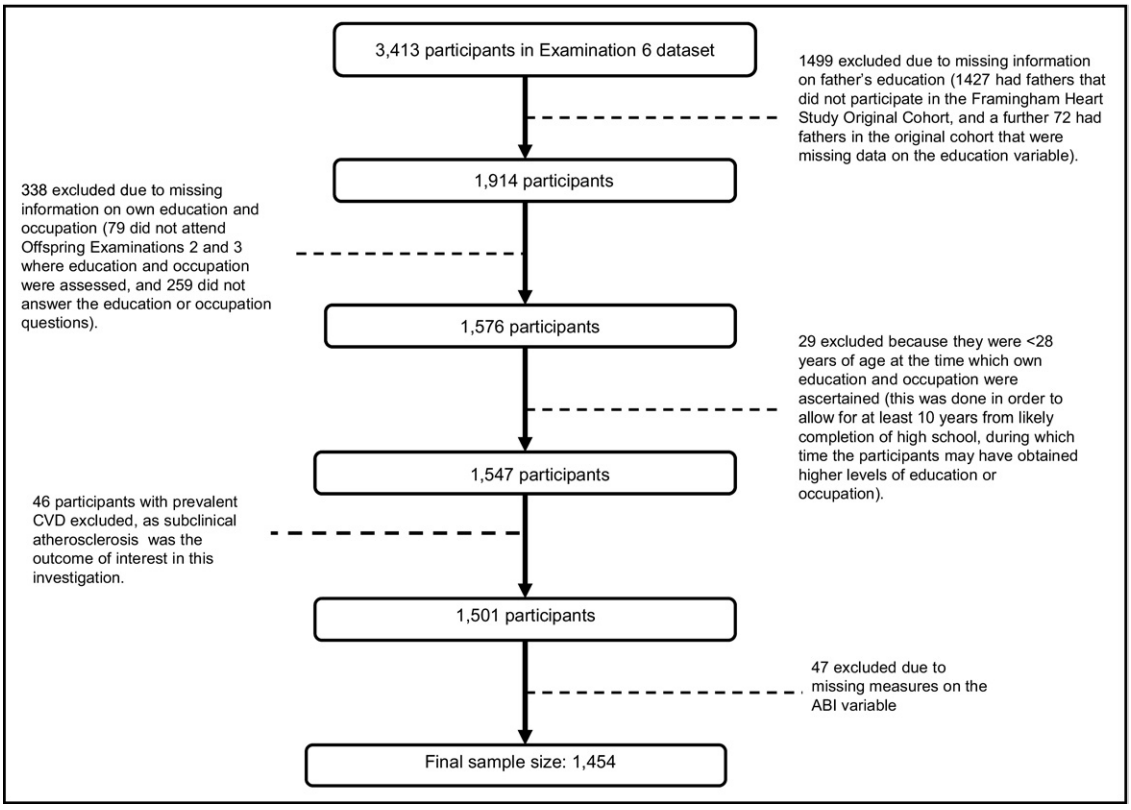


Figure 1. Exclusion criteria and final sample size for the present investigation from the Framingham Heart Offspring Study cohort in the United States (1971 to 1998). CVD = cardiovascular disease.

els: laborer, clerical/sales, and professional/executive/supervisory/technical. Young adulthood SEP was measured by own educational attainment, obtained directly from the Framingham Offspring Study participants at examinations 2 (1979 to 1982) and 3 (1984 to 1987). Education was categorized as 3 levels: ≤12, 13 to 16, and ≥17 years of education. Active professional-life SEP was measured as own occupation, ascertained at examination 2 (1979 to 1982), and categorized as 3 levels: laborer, homemaker/clerical/sales, and professional/executive/supervisory/technical. Analyses testing the accumulation-of-risk framework used a cumulative SEP score that was created by summing values for SEP at 3 successive life-course periods: childhood SEP (measured as father's education: lower than high school = 0, high school = 1, higher than high school = 2), young adulthood SEP (measured as own education: ≤12 years = 0, 13 to 16 years = 1, ≥17 years = 2), and active professional-life SEP (measured as own occupation: laborer = 0, clerical/sales/homemaker = 1, executive/professional/supervisory/technical = 2). Cumulative SEP score was categorized as low (score of 0 or 1), medium (score of 1 or 2), and high (score of 4 to 6) for analyses. Higher cutpoints were used for educational categories of offspring compared to fathers to account for secular trends of increased normative levels of education across generations.

Measurements of ABI were obtained at offspring examination 6 (1995 to 1998). Ankle–brachial systolic blood pressure measurements were performed by trained technicians according to standardized protocols.[4] Systolic blood pressure was measured using an 8-MHz Doppler pen probe and an ultrasonic Doppler flow detector (Parks Medical Electronics, Inc., Aloha, Oregon) 2 times for each limb (right arm, left arm, right ankle, left ankle). ABI was then calculated for each leg as the ratio of average systolic blood pressure in the ankle divided by average systolic blood pressure in the arm with the higher blood pressure. The lower of the ABI values calculated for the left and right ankles was used for analyses. If ABI was missing for 1 ankle, data from the nonmissing ankle were used (n = 8 for the present investigation).

Emerging evidence has demonstrated increased risk for coronary and carotid atherosclerosis, coronary events, and cardiovascular disease mortality with ABI values up to 1.1.[5–7] Consequently the ABI Collaboration defined normal ABI as >1.1 to 1.4.[5] Therefore, ABI was dichotomized as low (ABI ≤1.1) versus normal (ABI >1.1 to 1.4) for the present investigation. However, additional analyses were carried out for women using a lower cutpoint of 1.0 to define low ABI. This was done in light of recent evidence suggesting that normal ABI values may be intrinsically lower in healthy women than in men.[8] Participants with ABI values >1.4 were excluded because an ABI >1.4 has been demonstrated to confer increased risk for cardiovascular and total mortality, possibly because of poor arterial compressibility resulting from stiffness and calcification.[5] Because of a very small number of subjects (n = 41) with an ABI value ≤0.9 (i.e., definite peripheral arterial disease), there was inadequate statistical power to carry out analyses with ABI dichotomized as ≤0.9 versus >0.9 to 1.4.

Table 1
Age-adjusted characteristics according to cumulative life course socioeconomic position, Framingham Heart Study Offspring cohort, United States (1971 to 1998)

	Cumulative SEP Score		
	0 or 1	2 or 3	4–6
Men	192 (28.6%)	189 (28.1%)	291 (43.3%)
Age (years)*	58.9 (57.63–60.2)	56.3 (55.1–57.6)	55.6 (54.7–56.5)
Mean ankle–brachial index	1.15 (1.14–1.16)	1.15 (1.14–1.17)	1.17 (1.16–1.19)
Body mass index (kg/m^2)	29.0 (28.3–29.6)	29.0 (28.3–29.6)	28.2 (27.7–28.8)
Systolic blood pressure (mm Hg)	128.1 (125.9–130.3)	129.6 (127.4–131.7)	128.5 (126.8–130.3)
High-density lipoprotein cholesterol (mg/dl)	42.9 (41.1–44.8)	43.1 (41.3–44.9)	45.3 (43.8–46.8)
Total/high-density lipoprotein cholesterol ratio	5.1 (4.8–5.4)	5.4 (5.1–5.8)	4.7 (4.4–5.0)
Antihypertensive medication use	23.4% (17.8–30.0)	26.9% (20.9–34.0)	22.1% (17.6–27.5)
Cholesterol-lowering medication	11.4% (7.6–16.6)	12.3% (8.3–17.8)	10.4% (7.3–14.5)
Diabetes mellitus	9.4% (6.0–14.4)	7.2% (4.3–11.8)	8.4% (5.7–12.3)
Current smoker	15.4% (10.9–21.4)	19.5% (14.4–25.9)	10.0% (7.0–13.9)
Depression score	5.3 (4.3–6.2)	5.3 (4.4–6.3)	3.9 (3.1–4.7)
Women	230 (29.4%)	318 (40.7%)	234 (29.9%)
Age (years)*	61.1 (60.0–62.3)	56.1 (55.1–57.0)	55.9 (54.8–57.1)
Mean ankle–brachial index	1.08 (1.07–1.10)	1.09 (1.08–1.11)	1.09 (1.08–1.11)
Body mass index (kg/m^2)	27.9 (27.1–28.6)	27.6 (27.0–28.3)	26.4 (25.7–27.2)
Systolic blood pressure (mm Hg)	126.1 (123.8–128.5)	126.4 (124.4–128.3)	124.0 (121.8–126.3)
High-density lipoprotein cholesterol (mg/dl)	56.8 (54.7–58.9)	57.5 (55.7–59.2)	60.9 (58.9–62.9)
Total/high-density lipoprotein cholesterol ratio	4.1 (3.9–4.2)	4.0 (3.9–4.2)	3.7 (3.6–3.9)
Antihypertensive medication use	23.9% (18.7–30.0)	22.3% (17.9–27.4)	14.9% (10.8–20.2)
Cholesterol-lowering medication	8.8% (5.8–13.2)	6.2% (4.0–9.5)	6.4% (3.9–10.4)
Diabetes mellitus	7.9% (5.0–12.1)	7.4% (4.9–10.9)	4.5% (2.5–8.1)
Current smoker	25.2% (19.8–31.5)	16.2% (12.6–20.7)	11.5% (8.0–16.2)
Depression score	8.5 (7.4–9.6)	6.5 (5.6–7.5)	5.2 (4.1–6.3)

Data are expressed as mean or percent prevalence (95% confidence interval).
* Calculated using bivariate analysis.

All covariates were measured at offspring examination 6 (1995 to 1998). Smoking status (current, former, or never) was determined by self-report. Systolic blood pressure was calculated as the average of the clinic physician's 2 measurements of systolic blood pressure while a subject was seated. Body mass index was calculated as weight in kilograms divided by the square of height in meters. High-density lipoprotein (HDL) and total cholesterol concentrations were measured by automated enzymatic techniques.[9] Participants were considered to have diabetes if they reported receiving treatment with insulin or a hypoglycemic agent or if they had fasting plasma glucose levels ≥126 mg/dl (7.0 mmol/L). Participants who were missing information on treatment status or fasting glucose were classified as having diabetes if they had a nonfasting glucose concentration ≥200 mg/dl (11.1 mmol/L), or a nonfasting glucose concentration from 126 to 200 mg/dl plus a history of diabetes, or if nonfasting glucose concentration was 126 to 200 mg/dl with a diabetes diagnosis at a subsequent Framingham examination. Use of antihypertensive and cholesterol-lowering medication was self-reported. Depressive symptomatology was measured using the Center for Epidemiologic Studies Depression scale (range 0 to 51).

There were 3,413 participants in the dataset who completed offspring examination 6, on which the present investigation was based. After implementation of exclusion criteria (details shown in Figure 1), the final sample was 1,454 (782 women and 672 men). Cardiovascular disease events (as part of exclusion criteria) were identified in participants since the onset of the Framingham Offspring Study (1971 to 1975) and included recognized myocardial infarction, coronary insufficiency, cerebrovascular events (including cerebral atherothrombotic infarction, cerebral embolism, intracerebral hemorrhage, subarachnoid hemorrhage, and other cerebrovascular accident), and congestive heart failure. Those excluded (n = 1,913) were more likely to be older (mean age 60.0 vs 57.2 years, respectively, $p < 0.0001$), to be taking antihypertensive medication (31.6% vs 23.5%, $p < 0.0001$) and cholesterol-lowering medication (15.2% vs 10.0%, $p < 0.0001$), and to be diabetic (11.3% vs 8.4%, $p = 0.006$). Included and excluded participants did not differ significantly for other variables including gender, body mass index, HDL/total cholesterol ratio, depression score, and current smoking.

Age-adjusted means and proportions were calculated for baseline covariates and compared across cumulative SEP categories. Multivariable logistic regression analyses evaluated associations between cumulative SEP (categorized as low [score of 0 or 1], medium [score of 1 or 2], and high [score of 4 to 6] as described earlier) and ABI (≤1.1 vs >1.1 to 1.4 for primary analyses, ≤1.0 vs >1.0 to 1.4 for secondary analyses in women). All analyses were adjusted for age. Subsequent models were also adjusted for the CHD risk factors smoking, body mass index, systolic blood pressure, total/HDL cholesterol ratio, antihypertensive medication, cholesterol-lowering medication, depressive symptomatology, and diabetes. Further analyses of individual SEP measurements (father's education, own education, own oc-

Table 2
Odds ratios for association between life-course socioeconomic position and low ankle–brachial index defined as ankle–brachial index ≤1.10, Framingham Heart Study Offspring cohort, United States (1971 to 1998)

Cumulative SEP Score	Subjects	Number of Events With ABI ≤1.1 (%)	Model Adjustment			
			Model 1*		Model 2†	
			OR	95% CI	OR	95% CI
Men						
0 or 1	192	52 (27.1%)	2.00	1.28–3.14	2.04	1.22–3.42
2 or 3	189	45 (23.8%)	1.81	1.13–2.92	1.51	0.87–2.62
4–6	291	42 (14.4%)	1.00		1.00	
Women						
0 or 1	230	117 (50.9%)	0.94	0.63–1.38	0.86	0.56–1.33
2 or 3	318	153 (48.1%)	0.93	0.67–1.31	0.94	0.65–1.36
4–6	234	117 (50.0%)	1.00		1.00	

* Adjusted for age.
† Adjusted for age, smoking, body mass index, systolic blood pressure, total/high-density lipoprotein cholesterol ratio, antihypertensive medication, cholesterol-lowering medication, depressive symptomatology, and diabetes.

Table 3
Men—odds ratios for associations between socioeconomic position and low ankle–brachial index defined as ankle–brachial index ≤1.10, Framingham Heart Study Offspring cohort, United States (1971 to 1998)

SEP Measurement	SEP Level	Subjects	Number of Events With ABI ≤1.1 (%)	Model Adjustment					
				Model 1*		Model 2†		Model 3‡	
				OR	95% CI	OR	95% CI	OR	95% CI
Father's education	Lower than high school	331	72 (21.8%)	1.07	0.69–1.66	0.75	0.45–1.23	0.67	0.38–1.19
	High school	157	32 (20.4%)	1.2	0.71–2.04	0.96	0.54–1.71	1.00	0.52–1.92
	Higher than high school	184	35 (19.0%)	1.00		1.00		1.00	
Own education	≤12 years	216	59 (27.3%)	4.82	2.57–9.05	5.82	2.86–11.83	4.13	1.86–9.16
	13–16 years	270	67 (24.8%)	4.53	2.45–8.38	4.59	2.44–8.64	3.28	1.64–6.55
	≥17 years	186	13 (7.0%)	1.00		1.00		1.00	
Own occupation	Laborer	234	57 (24.3%)	1.55	1.02–2.35	0.92	0.57–1.50	1.22	0.70–2.11
	Homemaker, clerical, or sales	84	24 (28.6%)	1.91	1.10–3.32	1.40	0.79–2.51	1.92	1.01–3.62
	Professional, executive, supervisory, or technical	354	58 (16.4%)	1.00		1.00		1.00	

* Adjusted for age.
† Adjusted for age and other socioeconomic position measurements (i.e., socioeconomic position other than the exposure of interest; e.g., analyses on father's education are adjusted for own education and own occupation).
‡ Adjusted for age, other socioeconomic position measurements, smoking, body mass index, systolic blood pressure, total/high-density lipoprotein cholesterol ratio, antihypertensive medication, cholesterol-lowering medication, depressive symptomatology, and diabetes.

cupation) in relation to ABI were performed to evaluate whether SEP at any of these life-course periods particularly contributed to associations of cumulative SEP with ABI. These latter analyses were adjusted for age with subsequent adjustment for other individual SEP measurements and final adjustment for the CHD risk factors described earlier. All analyses performed were gender-specific because there was evidence of effect modification by gender ($p = 0.01$ for interaction between cumulative SEP score and gender). Generalized estimating equations were used to account for clustering of outcomes by family. Pearson correlation coefficients and variance inflation factors were used to evaluate colinearity, and no evidence of colinearity was found. The 3 primary SEP variables (father's education, own education, and own occupation) were not correlated highly enough to be of concern when simultaneously including all 3 in a single multivariable model (pairwise correlation coefficients ranged from 0.25 to 0.51). Power analyses were performed using PS Power and Sample Size Calculation 3.0.2 (Dupont and Plummer, Vanderbilt University, Nashville, Tennessee) according to methods for cohort studies with dichotomous outcomes.[10,11]

Results

Participants in the study sample had a mean age of 57 years (range 38 to 80) and 53.8% were women. Mean ABIs ± SDs were 1.16 ± 0.10 in men and 1.09 ± 0.10 in women ($p < 0.0001$). Prevalences of low ABI defined as ≤1.1 were 21% in men and 49% in women; low ABI defined as ≤1.0 demonstrated prevalences of 4.8% in men and 16% in women. In age-adjusted analyses in men lower cumulative SEP was associated with older age, lower ABI, higher body mass index, lower HDL cholesterol, higher total/HDL cholesterol ratio, higher prevalence of current smoking, and higher depression score. In women lower cumulative SEP was associated with older age, higher body mass index,

Table 4
Women—odds ratios for associations between socioeconomic position and low ankle–brachial index defined as ankle–brachial index ≤1.10, Framingham Heart Study Offspring cohort, United States (1971 to 1998)

SEP Measurement	SEP Level	Subjects	Number of Events With ABI ≤1.1 (%)	Model Adjustment					
				Model 1*		Model 2†		Model 3‡	
				OR	95% CI	OR	95% CI	OR	95% CI
Father's education	Lower than high school	388	185 (47.7%)	0.71	0.49–1.01	0.62	0.43–0.91	0.65	0.43–0.98
	High school	203	99 (48.8%)	0.86	0.58–1.29	0.85	0.56–1.28	0.94	0.60–1.47
	Higher than high school	191	103 (53.9%)	1.00		1.00		1.00	
Own education	≤12 years	333	173 (52.0%)	1.23	0.76–2.00	1.19	0.68–2.11	1.12	0.62–2.03
	13–16 years	355	172 (48.5%)	1.11	0.70–1.78	1.05	0.63–1.75	1.00	0.60–1.68
	≥17 years	94	42 (44.7%)	1.00		1.00		1.00	
Own occupation	Laborer	65	34 (52.3%)	1.36	0.76–2.41	1.42	0.75–2.68	1.12	0.56–2.23
	Homemaker, clerical, or sales	532	275 (51.7%)	1.40	1.00–1.96	1.40	0.79–2.51	1.31	0.87–1.98
	Professional, executive, supervisory, or technical	185	78 (42.2%)	1.00		1.00		1.00	

* Adjusted for age.

† Adjusted for age and other socioeconomic position measurements (i.e., socioeconomic position other than the exposure of interest; e.g., analyses on father's education are adjusted for own education and own occupation).

‡ Adjusted for age, other socioeconomic position measurements, smoking, body mass index, systolic blood pressure, total/high-density lipoprotein cholesterol ratio, antihypertensive medication, cholesterol-lowering medication, depressive symptomatology, and diabetes.

lower HDL cholesterol, higher total/HDL cholesterol ratio, higher prevalence of diabetes and current smoking, and higher depression score (Table 1). In addition, R^2 from univariate linear regression analyses showed that cumulative SEP explained 1.4% of the variance in ABI in men and 0.7% of the variance in women. Examples of the contribution of other specific CHD risk factors to the variance (R^2) in ABI were smoking (5.5% in men, 3.0% in women), hypertension medication (2.9% in men, 1.8% in women), systolic blood pressure (1.5% in men, 4.1% in women), and diabetes (1.3% in men, 0.1% in women).

Age-adjusted logistic regression analyses showed that lower cumulative SEP across the life course was associated with greater prevalence of low ABI in men (odds ratio [OR] 2.00, 95% confidence interval [CI] 1.28 to 3.14, for low vs high cumulative SEP score) and not in women (OR 0.94, 95% CI 0.63 to 1.38; Table 2). Further adjustment for CHD risk factors did not attenuate the association in men (OR 2.04, 95% CI 1.22 to 3.42). In analyses of individual SEP measurements own education was associated with low ABI in men with associations remaining after adjustment for age, other SEP measurements, and CHD risk factors (OR 4.13, 95% CI 1.86 to 9.16, for ≤12 years of education vs ≥17 years; Table 3). Own occupation was also associated with low ABI in men in age-adjusted analyses. Further adjustment for other SEP measurements and CHD risk factors attenuated the association for the laborer category (OR 1.22, 95% CI 0.70 to 2.11) but not for the homemaker/clerical/sales category (OR 1.92, 95% CI 1.01 to 3.62). No significant associations were observed between father's education and ABI in men (Table 3). In women no association was observed between own education and ABI (OR 1.23, 95% CI 0.76 to 2.00) or between own occupation and ABI (OR 1.36, 95% CI 0.76 to 2.41; Table 4). However, father's education in the lower-than-high-school category was weakly associated with a lower prevalence of low ABI after adjustment for other SEP measurements and CHD risk factors (Table 4). Additional analyses with low ABI alternatively defined as ≤1.0 in women revealed somewhat higher effect sizes; however, associations were still not statistically significant (supplementary Tables 1 and 2, available online). Patterns of association for cumulative SEP when using father's occupation as the measurement of childhood SEP were similar to results obtained when father's education was used (supplementary Table 3, available online). In addition, associations between father's occupation and ABI were not significant in men or women (data not shown), similar to analyses using father's education.

Discussion

Life-course cumulative SEP was inversely associated with low ABI, an indicator of subclinical atherosclerosis, in men and not in women. The effect in men appeared to be largely due to early adulthood SEP measured as participants' education compared to childhood SEP or active professional-life SEP. Adjustment for CHD risk factors did not attenuate the association in men, suggesting these may not be explanatory mechanisms. Few studies have investigated associations of cumulative SEP in relation to ABI. Carson et al[12] reported inverse associations between cumulative SEP and peripheral arterial disease (ABI <0.9) in middle-aged white men (n = 4,284) and women (n = 5,170) of the ARIC. Similar to our findings, they found that SEP in the young adulthood period was associated more strongly with peripheral arterial disease than in the childhood or older adulthood periods. Furthermore, associations of SEP with peripheral arterial disease were found in men and women, whereas our study did not find associations in women.

In previous publications evaluating relations between adulthood SEP only (rather than cumulative SEP) and ABI, studies in developed countries have shown generally similar findings to our study. Fowkes et al[13] demonstrated in a cross-sectional survey of participants 55 to 74 years old in the Edinburgh Artery Study that mean ABI decreased consistently with decreasing educational level in men and not in

women. Rooks et al[14] found inverse associations between various measurements of adulthood SEP (education, income, home ownership, and financial assets) and ABI in an elderly population of black and white men and women (n = 3,075). However, after adjustment for race, age, household family size, marital status, and study site, associations between education and low ABI persisted in men but not in women. Another study of 1,025 subjects in the Chianti area of Italy reported significantly lower age-adjusted mean ABIs in men but not in women with low education versus high education.[15]

Because there were large SEP gradients in some of the strongest risk factors for peripheral arterial disease (smoking and diabetes), lack of associations in women was somewhat surprising although consistent with other studies focused on associations of adulthood SEP with ABI.[13–15] One potential explanation is that the use of inappropriate cutpoints for defining low ABI in women may have contributed to null findings. There is evidence that ABI values are intrinsically lower in women compared to men[8] even after adjustment for height (which is suggested to contribute to the lower ABI observed in women).[13] McDermott et al[6] found evidence of excess coronary and carotid atherosclerosis with an ABI value up to 1.1 in men but only with an ABI value up to 1.0 in women. In our analyses, using lower ABI cutpoints ($\leq$1.0 instead of $\leq$1.1) to define low ABI somewhat increased effect sizes for associations between SEP and ABI in women; however, these effects were small and not statistically significant.

In our investigation measurements of childhood SEP (father's education and father's occupation) were not associated with low ABI with the exception of a slight positive association between father's education and low ABI in women. To the best of our knowledge the association between childhood SEP and subclinical atherosclerosis measured as ABI has rarely been examined. However, inverse associations between childhood SEP and clinically manifest CHD are more established and consistently observed.[16,17] In the present investigation those who were excluded from analyses because of cardiovascular disease were more likely to have lower childhood SEP and low ABI (data not shown) compared to those included. This may have led to an underestimation of an association between childhood SEP and low ABI in the study sample.

There is evidence suggesting that the effect of SEP on subclinical and clinically manifest CHD is partly mediated through CHD risk factors.[18–21] In this investigation adjustment for CHD risk factors did not attenuate the associations observed, similar to results from several previous studies reporting on associations between SEP and subclinical atherosclerosis.[22–25] Carson et al[12] reported that none of the risk factors tested in their study was a strong or moderate mediator of the association between SEP and peripheral arterial disease (ABI $<$0.9). However, it is important to note that methodologic biases may arise because of statistical adjustment for potential mediators and confounders.[26] Other potential risk factors not accounted for in this study may also play a role in the association between SEP and atherosclerosis development such as novel CHD risk markers,[27] psychosocial stressors,[28] and genetic susceptibility.[29]

A strength of this investigation was that measurements of childhood SEP were assessed directly from the participants' parents. Thus this measurement was less likely to be subject to measurement error compared to studies that obtained measurements of childhood SEP retrospectively through personal recall by participants.[16] Furthermore, rigorous quality assurance and quality control methods were used in this study to ensure high-quality measurements of outcomes and covariates.

Several limitations of this study should be noted. Subjects in this study population were predominantly of European descent (representing the demographics of the city of Framingham, Massachusetts at study onset); thus results from this study are not necessarily generalizable to other races and ethnicities. Furthermore, primary analyses on life-course SEP included only 3 measurements: father's education, own education, and own occupation. Additional measurements (such as SEP during in utero period) or more comprehensive measurements (such as multiple measurements of occupation throughout the life course) of life-course SEP would provide richer data on the potential contributions of life-course SEP to low ABI.

Acknowledgment: The authors are grateful to the National Heart, Lung, and Blood Institute for providing them with the limited access dataset in conjunction with Framingham Heart Study investigators who provided other important variables for analyses.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.amjcard.2011.07.030.

1. Weatherley BD, Nelson JJ, Heiss G, Chambless LE, Sharrett AR, Nieto FJ, Folsom AR, Rosamond WD. The association of the ankle-brachial index with incident coronary heart disease: the Atherosclerosis Risk In Communities (ARIC) study, 1987–2001. *BMC Cardiovasc Disord* 2007;7:3–13.
2. Heald CL, Fowkes FG, Murray GD, Price JF; Ankle Brachial Index Collaboration. Risk of mortality and cardiovascular disease associated with the ankle–brachial index: systematic review. *Atherosclerosis* 2006;189:61–69.
3. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham Offspring Study. *Am J Epidemiol* 1979;110:281–290.
4. Murabito JM, Guo CY, Fox CS, D'Agostino RB. Heritability of the ankle–brachial index: the Framingham Offspring Study. *Am J Epidemiol* 2006;164:963–968.
5. Fowkes FG, Murray GD, Butcher I, Heald CL, Lee RJ, Chambless LE, Folsom AR, Hirsch AT, Dramaix M, deBacker G, Wautrecht JC, Kornitzer M, Newman AB, Cushman M, Sutton-Tyrrell K, Lee AJ, Price JF, d'Agostino RB, Murabito JM, Norman PE, Jamrozik K, Curb JD, Masaki KH, Rodriguez BL, Dekker JM, Bouter LM, Heine RJ, Nijpels G, Stehouwer CD, Ferrucci L, McDermott MM, Stoffers HE, Hooi JD, Knottnerus JA, Ogren M, Hedblad B, Witteman JC, Breteler MM, Hunink MG, Hofman A, Criqui MH, Langer RD, Fronek A, Hiatt WR, Hamman R, Resnick HE, Guralnik J, Resnick HE, Guralnik J, McDermott MM. Ankle brachial index combined with Framingham Risk Score to predict cardiovascular events and mortality: a meta-analysis. *JAMA* 2008;300:197–208.
6. McDermott MM, Liu K, Criqui MH, Ruth K, Goff D, Saad MF, Wu C, Homma S, Sharrett AR. Ankle–brachial index and subclinical cardiac and carotid disease: the multi-ethnic study of atherosclerosis. *Am J Epidemiol* 2005;162:33–41.
7. O'Hare AM, Katz R, Shlipak MG, Cushman M, Newman AB. Mortality and cardiovascular risk across the ankle-arm index spectrum:

results from the Cardiovascular Health Study. *Circulation* 2006;113: 388–393.
8. Aboyans V, Criqui MH, McClelland RL, Allison MA, McDermott MM, Goff DC Jr, Manolio TA. Intrinsic contribution of gender and ethnicity to normal ankle-brachial index values: the Multi-Ethnic Study of Atherosclerosis (MESA). *J Vasc Surg* 2007;45:319–327.
9. McNamara JR, Schaefer EJ. Automated enzymatic standardized lipid analyses for plasma and lipoprotein fractions. *Clin Chim Acta* 1987; 166:1–8.
10. Casagrande JT, Pike MC. An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics* 1978;34:483–486.
11. Fleiss JL. Statistical Methods for Rates and Proportions. New York: John Wiley, 1981:38–46.
12. Carson AP, Rose KM, Catellier DJ, Kaufman JS, Wyatt SB, Diez-Roux AV, Heiss G. Cumulative socioeconomic status across the life course and subclinical atherosclerosis. *Ann Epidemiol* 2007;17: 296–303.
13. Fowkes FG, Housley E, Cawood EH, Macintyre CC, Ruckley CV, Prescott RJ. Edinburgh Artery Study: prevalence of asymptomatic and symptomatic peripheral arterial disease in the general population. *Int J Epidemiol* 1991;20:384–392.
14. Rooks RN, Simonsick EM, Miles T, Newman A, Kritchevsky SB, Schulz R, Harris T. The association of race and socioeconomic status with cardiovascular disease indicators among older adults in the health, aging, and body composition study. *J Gerontol B Psychol Sci Soc Sci* 2002;57:S247–S256.
15. Coppin AK, Ferrucci L, Lauretani F, Phillips C, Chang M, Bandinelli S, Guralnik JM. Low socioeconomic status and disability in old age: evidence from the InChianti study for the mediating role of physiological impairments. *J Gerontol A Biol Sci Med Sci* 2006;61:86–91.
16. Galobardes B, Smith GD, Lynch JW. Systematic review of the influence of childhood socioeconomic circumstances on risk for cardiovascular disease in adulthood. *Ann Epidemiol* 2006;16:91–104.
17. Loucks EB, Lynch JW, Pilote L, Fuhrer R, Almeida ND, Richard H, Agha G, Murabito JM, Benjamin EJ. Life-course socioeconomic position and incidence of coronary heart disease: the Framingham Offspring Study. *Am J Epidemiol* 2009;169:829–836.
18. Dragano N, Verde PE, Moebus S, Stang A, Schmermund A, Roggenbuck U, Möhlenkamp S, Peter R, Jöckel KH, Erbel R, Siegrist J; Heinz Nixdorf Recall Study. Subclinical coronary atherosclerosis is more pronounced in men and women with lower socio-economic status: associations in a population-based study. Coronary atherosclerosis and social status. *Eur J Cardiovasc Prev Rehabil* 2007;14:568–574.
19. Harald K, Pajunen P, Jousilahti P, Koskinen S, Vartiainen E, Salomaa V. Modifiable risk factors have an impact on socio-economic differences in coronary heart disease events. *Scand Cardiovasc J* 2006;40: 87–95.
20. Nordstrom CK, Diez Roux AV, Jackson SA, Gardin JM; Cardiovascular Health Study. The association of personal and neighborhood socioeconomic indicators with subclinical cardiovascular disease in an elderly cohort. The Cardiovascular Health Study. *Soc Sci Med* 2004; 59:2139–2147.
21. Yan LL, Liu K, Daviglus ML, Colangelo LA, Kiefe CI, Sidney S, Matthews KA, Greenland P. Education, 15-year risk factor progression, and coronary artery calcium in young adulthood and early middle age: the Coronary Artery Risk Development in Young Adults study. *JAMA* 2006;295:1793–1800.
22. Gallo LC, Matthews KA, Kuller LH, Sutton-Tyrrell K, Edmundowicz D. Educational attainment and coronary and aortic calcification in postmenopausal women. *Psychosom Med* 2001;63:925–935.
23. Lutsey PL, Diez Roux AV, Jacobs DR Jr, Burke GL, Harman J, Shea S, Folsom AR. Associations of acculturation and socioeconomic status with subclinical cardiovascular disease in the multi-ethnic study of atherosclerosis. *Am J Public Health* 2008;98:1963–1970.
24. Lynch J, Kaplan GA, Salonen R, Salonen JT. Socioeconomic status and progression of carotid atherosclerosis. Prospective evidence from the Kuopio Ischemic Heart Disease Risk Factor Study. *Arterioscler Thromb Vasc Biol* 1997;17:513–519.
25. Ranjit N, Diez-Roux AV, Chambless L, Jacobs DR Jr, Nieto FJ, Szklo M. Socioeconomic differences in progression of carotid intima-media thickness in the Atherosclerosis Risk in Communities study. *Arterioscler Thromb Vasc Biol* 2006;26:411–416.
26. Cole SR, Hernán MA. Fallibility in estimating direct effects. *Int J Epidemiol* 2002;31:163–165.
27. Allison MA, Criqui MH, McClelland RL, Scott JM, McDermott MM, Liu K, Folsom AR, Bertoni AG, Sharrett AR, Homma S, Kori S. The effect of novel cardiovascular risk factors on the ethnic-specific odds for peripheral arterial disease in the Multi-Ethnic Study of Atherosclerosis (MESA). *J Am Coll Cardiol* 2006;48:1190–1197.
28. Lynch J, Smith GD. A life course approach to chronic disease epidemiology. *Annu Rev Public Health* 2005;26:1–35.
29. Trichopoulou A, Yiannakouris N, Bamia C, Benetou V, Trichopoulos D, Ordovas JM. Genetic predisposition, nongenetic risk factors, and coronary infarct. *Arch Intern Med* 2008;168:891–896.

HEALTH CARE REFORM

Collaborative Care of Opioid-Addicted Patients in Primary Care Using Buprenorphine

Five-Year Experience

Daniel P. Alford, MD, MPH; Colleen T. LaBelle, RN; Natalie Kretsch, BA; Alexis Bergeron, MPH, LCSW; Michael Winter, MPH; Michael Botticelli, MEd; Jeffrey H. Samet, MD, MA, MPH

Background: Opioid addiction is a chronic disease treatable in primary care settings with buprenorphine hydrochloride, but this treatment remains underused. We describe a collaborative care model for managing opioid addiction with buprenorphine hydrochloride–naloxone hydrochloride dihydrate sublingual tablets.

Methods: Ours is a cohort study of patients treated for opioid addiction using collaborative care between nurse care managers and generalist physicians in an urban academic primary care practice during a 5-year period. We examine patient characteristics, 12-month treatment success (ie, retention or taper after 6 months), and predictors of successful outcomes.

Results: From September 1, 2003, through September 30, 2008, 408 patients with opioid addiction were treated with buprenorphine. Twenty-six patients were excluded from analysis because they left treatment owing to preexisting legal or medical conditions or a need to transfer to another buprenorphine program. At 1 year, 196 of 382 patients (51.3%) underwent successful treatment. Of patients remaining in treatment at 12 months, 154 of 169 (91.1%) were no longer using illicit opioids or cocaine based on urine drug test results. On admission, patients who were older, were employed, and used illicit buprenorphine had significantly higher odds of treatment success; those of African American or Hispanic/Latino race had significantly lower odds of treatment success. These outcomes were achieved with a model that facilitated physician involvement.

Conclusion: Collaborative care with nurse care managers in an urban primary care practice is an alternative and successful treatment method for most patients with opioid addiction that makes effective use of time for physicians who prescribe buprenorphine.

Arch Intern Med. 2011;171(5):425-431

Author Affiliations: Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center (Drs Alford and Samet and Mss LaBelle, Kretsch, and Bergeron), Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine (Drs Alford and Samet), Data Coordinating Center (Mr Winter) and Department of Community Health Sciences (Dr Samet), Boston University School of Public Health, and Bureau of Substance Abuse Services, Department of Public Health (Mr Botticelli), Boston, Massachusetts.

OPIOID ADDICTION IS A chronic, relapsing brain disease that affects millions of Americans and produces tremendous burden on the health care system.[1,2] Recent epidemiologic studies[3] have revealed alarming increases in opioid addiction and overdoses, particularly with regard to prescription opioids. Less than 25% of individuals addicted to opioids receive addiction treatment.[3] For more than 45 years, research has confirmed that opioid agonist therapy (ie, methadone hydrochloride) is a highly effective treatment for opioid addiction provided outside primary care.[4-6] In 2002, US physicians gained the opportunity to treat opioid-addicted patients with buprenorphine hydrochloride in primary care settings, commonly referred to as office-based opioid treatment (OBOT).[7] This treatment has been shown to be effective in primary care settings[8-15]; however, it remains underused in traditional care models.[16] One consistently cited barrier preventing OBOT expansion is lack of adequate clinical support given the additional needs for patient monitoring.[16-18] Although collaborative care improves management of other chronic diseases (eg, hypertension[19] and diabetes mellitus[20]), experience with this model for the treatment of opioid addiction in primary care has not been described.[21]

On September 1, 2003, an OBOT program using a collaborative care model was started at Boston Medical Center, an urban academic medical setting. This model accommodated the large public demand for OBOT and faculty physicians with part-time clinical practices. We describe Boston Medical Center's OBOT program and report on patient and program outcomes.

METHODS

PROGRAM DESCRIPTION

Collaborative Model of Care

The collaborative care OBOT program included a full-time nurse program director, nurse care managers (NCMs), a program coordinator, and generalist physicians with part-time clinical practices. The nurse program director (0.40 full-time equivalent) supervised the NCMs and program coordinator and integrated care with OBOT physicians. The program coordinator (1 full-time equivalent), a former medical assistant, was trained to collect standardized intake information regarding individuals requesting OBOT. The NCMs, registered nurses who completed a 1-day buprenorphine training program, performed patient care roles, followed treatment protocols, and maintained a standard of clinical practice that satisfied federal regulations for buprenorphine treatment. Their clinical responsibilities included assessing for the appropriateness for OBOT, educating patients, obtaining informed consent, developing treatment plans, overseeing medication management, referring to other addiction treatment, monitoring for treatment adherence, and communicating with prescribing physicians, addiction counselors, and pharmacists. Collaboration with pharmacists reduced the OBOT physician burden by allowing buprenorphine prescriptions with multiple refills while allowing for cancellation of the refills if the patient showed nonadherence. The OBOT program currently includes NCMs (2.2 full-time equivalents) for 22 clinical half-day sessions per week. The OBOT physicians, all generalists with part-time clinical practices, reviewed and supplemented the NCM assessments (including laboratory results), performed physical examinations, prescribed buprenorphine, and followed up patients at least every 6 months or more frequently if needed. The OBOT program includes 9 generalist physicians, all waivered to prescribe buprenorphine by completing the required 8 hours of buprenorphine training; 3 of them (including D.P.A. and J.H.S.) are certified by the American Board of Addiction Medicine. The physicians had an average of 3 primary care half-day sessions each week, ranging from 1 to 6.

The treatment model included 3 stages: (1) NCM and physician assessment (appropriateness for OBOT and intake evaluations), (2) NCM-supervised induction and stabilization (buprenorphine dose adjustments during days 1-7), and (3) maintenance (buprenorphine treatment with monitoring for illicit drug use and weekly counseling) or discharge (voluntary or involuntary).

Assessment

The program coordinator conducted a scripted screening, by telephone or in person, documenting substance use, addiction treatment history, medical and psychiatric history, medications, addiction treatment goals, and availability of social support. The nurse program director reviewed the screen and triaged the patient to NCM and physician intake appointments or to other treatment options. Patients were triaged to other treatment options (eg, detoxification program) if they had active co-occurring substance use disorders. The NCM intake included documentation of opioid dependence diagnosis based on a checklist of the *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition, Text Revision)[22] criteria, assessing ability to adhere to treatment plan, mental health stability, polysubstance use, absence of painful conditions requiring long-term opioid analgesic therapy, and absence of medical contraindications (eg, pregnancy). All patients transferring from methadone maintenance were required to taper their methadone dose to 30 mg/d.[23] Individuals were excluded if they were unwilling to stop all illicit drug use or were not interested in buprenorphine maintenance lasting at least 6 months. Patients were educated regarding the scientific basis of buprenorphine maintenance. All patients signed an informed consent and treatment agreement, which included a weekly counseling requirement with a release for communication. Initial laboratory tests included viral hepatitis and syphilis serologic tests, liver function tests, pregnancy tests, and a urine drug test for opiates, cocaine, benzodiazepines, barbiturates, and amphetamines. Because many of these patients were new to primary health care, laboratory tests for a broader primary care evaluation were also ordered, including a complete blood cell count and electrolyte levels. Starting June 1, 2006, all patients were also tested for oxycodone, methadone, and buprenorphine. Patients were required to test negative for all nonprescribed substances other than opioids before starting buprenorphine treatment.

The OBOT physician intake included reviewing and supplementing the NCM assessment and treatment plan, performing a physical examination, and evaluating any medical issues (eg, need for hepatitis vaccinations). Patients with active psychiatric diagnoses were comanaged with a psychiatrist, with releases to facilitate communication.

Induction and Stabilization

The physician prescribed buprenorphine based on federal guidelines.[23] Induction and stabilization occurred during the first 7 days. Induction occurred on site (ie, in the office) with direct observation by the NCM for signs of opioid withdrawal before the first dose of buprenorphine and signs of precipitated withdrawal after the first dose. Physicians were not present during buprenorphine induction but were available by pager for consultation. Dose adjustments were made based on OBOT program protocols. The combination tablet buprenorphine hydrochloride–naloxone hydrochloride dihydrate was used for all inductions. During dose stabilization (days 2-7), dosing occurred off site (ie, at home), with patients having at least daily telephone contact with the NCM to assess for withdrawal and adverse effects. Telephone support continued until a stable maintenance dose was achieved, usually 8 to 16 mg. All patients were reassessed in person by the NCM on day 8 or sooner if needed.

Maintenance

Ongoing monitoring occurred at follow-up appointments at least weekly for the first 4 to 6 weeks followed by visits every 2 to 4 weeks if adherent (ie, 4 consecutive urine drug tests with negative results for illicit drugs and positive results for buprenorphine and attending at least 3 of 4 counseling sessions per month). Patients who maintained sobriety and treatment adherence for 3 months had their NCM visit frequency decreased to a monthly basis, then once every 3 months. Patients seen on a less than monthly basis had up to 6 random callbacks per year for unscheduled urine drug testing and pill counts. Patients were always subject to requests to return to the clinic within 72 hours for urine drug tests, observed dosing, pill counts, or treatment plan revisions. Urine drug tests occurred at every NCM visit and OBOT physician visit at least every 6 months. All patients with abnormal urine drug test results were called in the following week to meet with the NCM. If a patient continued to use illicit drugs (eg, opiates, cocaine, and benzodiazepines) or showed nonadherence with scheduled appointments or monitoring requests (ie, urine drug tests or pill counts), the NCM intensified treatment, including increased visits to 2 or more times per week for pill

counts, observed urine drug tests and adjusted buprenorphine dose, reengaged social supports, and intensified counseling.

The NCMs and physicians encouraged patients to attend self-help groups and mandatory weekly addiction counseling. Most addiction counseling (individual and group) was offered by outside addiction services with confidentiality releases. For an 18-month period, a few patients were seen by an in-house counselor. The NCM reminded patients of upcoming appointments with addiction counselors and OBOT physicians.

Discharge

Patients were referred to methadone maintenance treatment if they continued to use illicit opiates as determined by 3 consecutive urine drug test results positive for opiates despite increased buprenorphine dose or if they needed more structured care (eg, daily observed dosing due to concerns regarding medication misuse) than could be offered in an office-based practice. Patients were involuntarily discharged if they declined transfer to methadone maintenance after continued illicit opiate use or showed repeated nonadherence with more than 3 OBOT appointments or monitoring requests (ie, supplying urine sample for drug testing or bringing in remaining pills for pill counts). Patients were also discharged if they engaged in disruptive behavior. Patients could request a buprenorphine taper at any time but were encouraged to wait until they had achieved at least 6 months of abstinence and treatment adherence.

PATIENT CHARACTERISTICS AND DATA COLLECTION PROCEDURE

Patient characteristics documented included age; sex; race; employment; history of psychiatric illness; human immunodeficiency virus and hepatitis status; opioid of choice; years of opioid use; history of overdose; other substance use (tobacco, alcohol, cocaine, or benzodiazepines); previous addiction treatment, including history of opioid agonist therapy (ie, methadone or buprenorphine); and current use of illicit buprenorphine. A retrospective clinical medical record review was conducted for all patients admitted to the Boston Medical Center OBOT program from September 1, 2003, through September 30, 2008, with 12-month outcome data continued through September 30, 2009. All data were originally collected by OBOT clinical staff for the purposes of clinical care.

OUTCOME ASSESSMENTS

Treatment Outcomes at 12 Months

Treatment status (ie, successful, unsuccessful, or methadone maintenance transfer) was determined for all patients at 12 months or at program departure, whichever came first. Successful treatment was defined as treatment retention or buprenorphine taper after treatment adherence and absence of illicit drug use for at least 6 months. Treatment retention required patients to demonstrate a consistent pattern of adherence, including physician, nursing, and counseling appointments; monitoring requests (ie, urine drug tests or pill counts); no buprenorphine diversion (ie, accurate pill counts or positive buprenorphine drug test results); and willingness to engage in intensified treatment (eg, increased frequency of NCM visits) when illicit drug use occurred. Unsuccessful treatment included loss to follow-up, involuntary discharge due to continued illicit drug use, treatment nonadherence or disruptive behavior, or voluntary discharge due to adverse effects of buprenorphine. Treatment was considered neither successful nor unsuccessful if patients voluntarily transferred to methadone maintenance treatment for more structured care (eg, observed daily dosing) or for full opioid agonist therapy due to opioid craving while taking maximum doses of buprenorphine. Records were reviewed for the specific date of and reason for discharge from the program. Patients who left the program because of preexisting legal or medical conditions or transferred to another OBOT program because of relocation or consolidation of their care (eg, the same psychiatrist to treat mental illness and opioid addiction) were not included in analyses because treatment discharge was not related to the buprenorphine treatment program.

Illicit Drug Use

Both scheduled and random callback urine drug tests were conducted at least once every 3 months. In each study assessment window (ie, 3, 6, 9, and 12 months), the result of the test performed the closest, yet prior, to the time point was examined. For those patients who missed appointments within the 3-month interval, no urine sample was obtained; consequently, the number of urine samples available at each testing interval was less than the number of patients enrolled in that interval. Patients who were unable to be contacted were still considered enrolled in the program until their 30-day prescriptions ran out. Urine drug tests were mostly unsupervised, but measures were taken to try to minimize falsified tests (eg, testing urine temperature). Urine collections were supervised for patients with a recent abnormal urine test result, including a cold or diluted specimen, and for patients with aberrant behaviors (eg, missed appointments).

Program Activity

The average workload for the program coordinator was determined by tracking inquiries for OBOT treatment. Caseload for NCMs and OBOT physicians was determined by their schedules.

STATISTICAL ANALYSIS

Descriptive statistics (eg, means and proportions) were used to characterize the sample. Exploratory, hypothesis-generating tests were then performed to compare characteristics between patients with successful and unsuccessful treatments at 12 months. The χ^2 or Fisher exact tests (for dichotomous variables) and t tests or Wilcoxon rank sum tests (for continuous variables) were used to assess factors associated with treatment success at 12 months. Characteristics that were significantly associated in bivariate analyses, along with other characteristics deemed clinically important, were entered into a multivariable logistic regression model, with treatment success as the dependent variable. Reported P values are 2-tailed, and $P < .05$ was considered statistically significant. All analyses were run using SAS/STAT software, version 9.1 (SAS Institute Inc, Chicago, Illinois). This research was approved by the institutional review board at Boston University Medical Center.

RESULTS

PATIENT CHARACTERISTICS

In 5 years, 408 patients with opioid dependence were admitted to the OBOT program. Twenty-six patients were excluded from exploratory analysis because they left treatment owing to preexisting legal issues leading to incarceration (n=13) and medical conditions unrelated to their buprenorphine treatment (chronic pain [n=3] and

Table 1. Characteristics of 382 Opioid-Dependent Patients Entering Office-Based Opioid Treatment in Primary Care

Characteristic	Total, No.	Patients, No. (%)[a] Successful Treatment (n=196 [51.3%])	Unsuccessful Treatment (n=162 [42.4%])	Methadone Hydrochloride Transfer (n=24 [6.3%])	*P* Value
Sex					
Male	252	120 (47.6)	117 (46.4)	15 (6.0)	.09
Female	130	76 (58.5)	45 (34.6)	9 (6.9)	
Race					
White	254	146 (57.5)	94 (37.0)	14 (5.5)	.03
Hispanic/Latino	59	20 (33.9)	34 (57.6)	5 (8.5)	
Black/African American	63	27 (42.9)	31 (49.2)	5 (7.9)	
Other/unspecified	6	3 (50.0)	3 (50.0)	0	
Age on admission, mean (SD), y	39 (11)	39 (11)	38 (11)	39 (12)	.29
Employed at time of admission					
Yes	132	84 (63.6)	41 (31.1)	7 (5.3)	.002
No	249	112 (45.0)	120 (48.2)	17 (6.8)	
Data Missing	1		1 (1.06)		
Psychiatric illness (self-reported)					
Yes	252	132 (52.4)	105 (41.7)	15 (6.0)	.82
No	130	64 (49.2)	57 (43.8)	9 (6.9)	
HIV infected (self-reported)					
Yes	57	27 (13.8)	26 (45.6)	4 (7.0)	.94
No	312	162 (82.6)	130 (41.7)	20 (6.4)	
Unknown	13	7 (3.6)	6 (46.2)	0	
Hepatitis C infected (laboratory report)					
Yes	192	90 (45.9)	92 (47.9)	10 (5.2)	.08
No	190	106 (54.1)	70 (36.8)	14 (7.4)	
Opioid use at time of admission					
Heroin	231	105 (45.5)	115 (49.8)	11 (4.8)	<.001
Prescription opioids only	64	41 (64.1)	21 (32.8)	2 (3.1)	
Methadone maintenance treatment	51	29 (56.9)	12 (23.5)	10 (19.6)	
Buprenorphine hydrochloride maintenance treatment	36	21 (58.3)	14 (38.9)	1 (2.8)	
Illicit buprenorphine use at time of admission					
Yes	40	27 (13.8)	9 (22.5)	4 (10.0)	.02
No	342	169 (86.2)	153 (44.7)	20 (5.8)	
Opioid use, median (range), y	12 (1-40)	12 (1-40)	10 (1-40)	14 (1-32)	.65
History of opioid overdose					
Yes	176	82 (41.8)	79 (44.9)	15 (8.5)	.11
No	205	113 (57.6)	83 (40.5)	9 (4.4)	
Data missing	1	1 (0.6)			
Tobacco use at time of admission (past year, self-reported)					
Yes	304	157 (80.1)	131 (43.1)	16 (5.3)	.24
No	77	39 (19.9)	30 (39.0)	8 (10.4)	
Data missing	1	1 (0.5)			
Alcohol use at time of admission (past year, self-reported)					
Yes	180	94 (48.0)	76 (42.2)	10 (5.6)	.84
No	202	102 (52.0)	86 (42.6)	14 (6.9)	
Cocaine use at time of admission (past year, self-reported)					
Yes	165	73 (37.2)	83 (50.3)	9 (5.4)	.02
No	217	123 (62.8)	79 (36.4)	15 (6.9)	
Any past opioid agonist maintenance therapy					
Yes	221	122 (62.2)	84 (38.0)	15 (6.8)	.13
No	161	74 (37.8)	78 (48.4)	9 (5.6)	
Methadone maintenance					
Yes	165	87 (44.4)	63 (38.2)	15 (9.1)	.08
No	217	109 (55.6)	99 (45.6)	9 (4.2)	
Buprenorphine maintenance					
Yes	82	48 (24.5)	31 (37.8)	3 (3.7)	.26
No	300	148 (75.5)	131 (43.7)	21 (7.0)	
Past inpatient detoxification					
Yes	323	157 (80.1)	145 (44.9)	21 (6.5)	.05
No	59	39 (19.9)	17 (28.8)	3 (5.1)	
Past residential treatment					
Yes	194	103 (52.5)	79 (40.7)	12 (6.2)	.77
No	188	93 (47.5)	83 (44.2)	12 (6.4)	

Abbreviation: HIV, human immunodeficiency virus.
[a]Data are presented as number (percentage) of patients unless otherwise indicated. Percentages may not total 100 because of rounding.

Table 2. Treatment Outcomes at 12 Months of 382 Opioid-Dependent Patients Entering Office-Based Opioid Treatment in Primary Care

Outcome	Patients, No. (%)
Successful treatment	196 (51.3)
Treatment retention	187 (49.0)
Successful taper after 6 months of adherence[a]	9 (2.4)
Unsuccessful treatment	162 (42.4)
Lost to follow-up	113 (29.6)
Nonadherence despite enhanced treatment[a]	46 (12.0)
Administrative discharge due to disruptive behavior	2 (0.5)
Adverse effects of buprenorphine hydrochloride	1 (0.3)
Transfer to methadone hydrochloride treatment program	24 (6.3)

[a] Adherence was defined as attending scheduled office-based opioid treatment program appointments, complying with required monitoring (ie, urine drug tests or pill counts), absence of evidence of buprenorphine diversion, and lack of sustained illicit opiate use.

advanced-stage AIDS [n=1]) or transferred to another OBOT program because of relocation or consolidation of their care (n=9), leaving 382 patients. This group, described in **Table 1**, was predominantly male (252 [66.0%]) and white (254 [66.5%]). The mean age was 39 years; 132 (34.6%) were employed at admission. Comorbidities were common; 252 (66.0%) reported psychiatric illness and 192 (50.3%) tested positive for hepatitis C antibody. On admission, patients were using the following: heroin (with or without prescription opioids), 231 (60.5%); prescription opioids exclusively, 64 (16.8%); methadone from a maintenance program, 51 (13.4%); and buprenorphine from another OBOT program, 36 (9.4%). Past-year use of tobacco (304 [79.6%]) and cocaine (165 [43.2%]) was common. A total of 323 (84.6%) reported a history of inpatient detoxification; 221 (57.9%), past opioid agonist maintenance treatment; 40 (10.5%), current use of illicit buprenorphine; and 176 (46.1%), history of opioid overdose.

PATIENT OUTCOMES

Patient outcomes are described in **Table 2**. At 12 months, 196 of the 382 patients (51.3%) showed a successful outcome (187 [49.0%] remained in treatment and 9 [2.4%] were tapered after 6 months of adherence and the absence of illicit drug use), and 162 patients (42.4%) showed an unsuccessful outcome (113 [29.6%] lost to follow-up and 49 [12.8%] discharged). A few patients (24 [6.3%]) had courses of treatment that were considered neither successful nor unsuccessful because those patients voluntarily transferred to methadone maintenance treatment for more structured care or for full opioid agonist therapy.

Patients who achieved successful outcomes were compared with those who had unsuccessful outcomes to determine factors associated with treatment success. In bivariate analyses, treatment success at 12 months was significantly associated with the following characteristics on admission: female sex, white race, being employed, self-maintaining with illicit buprenorphine, prescription opioid abuse or methadone maintenance, and no self-report of past-year cocaine use.

Table 3. Multivariable Logistic Regression for Factors That Influence Treatment Success at 12 Months for 356 Opioid-Dependent Patients Entering Office-Based Opioid Treatment in Primary Care[a]

Characteristic	OR (95% CI)
Sex	
Female	1.65 (0.97-2.81)
Male	1 [Reference]
Race	
African American	0.50 (0.26-0.99)[b]
Hispanic	0.45 (0.22-0.93)[b]
Other	0.37 (0.06-2.22)
White	1 [Reference]
Age, 10-y increase	1.40 (1.09-1.80)[c]
Employed	
Yes	2.24 (1.33-3.77)[c]
No	1 [Reference]
Hepatitis C infected (laboratory report)	
Yes	0.75 (0.43-1.29)
No	1 [Reference]
Opioid use on admission	
Prescribed buprenorphine hydrochloride	1.06 (0.45-2.48)
Methadone hydrochloride maintenance	2.02 (0.87-4.67)
Prescription opioids only	1.01 (0.48-2.12)
Heroin	1 [Reference]
Illicit buprenorphine use at time of admission	
Yes	3.04 (1.32-7.00)[c]
No	1 [Reference]
Cocaine use at time of admission (past year, self-reported)	
Yes	0.71 (0.44-1.15)
No	1 [Reference]
Previous history of detoxification admission	
Yes	0.59 (0.28-1.25)
No	1 [Reference]
Previous history of methadone maintenance treatment	
Yes	1.22 (0.70-2.11)
No	1 [Reference]
Psychiatric illness (self-reported)	
Yes	1.19 (0.72-1.99)
No	1 [Reference]
History of opioid overdose	
Yes	0.97 (0.58-1.63)
No	1 [Reference]

Abbreviations: CI, confidence interval; OR, odds ratio.
[a] Twenty-four patients transferred to methadone maintenance were excluded from this analysis, and data were missing for an additional 2 patients.
[b] $P<.05$.
[c] $P<.01$.

These characteristics, along with factors considered to be clinically important to successful treatment (ie, age, past methadone maintenance treatment, self-reported psychiatric illness, and history of opioid overdose), were entered into a multivariable logistic regression model (**Table 3**). The patients who were transferred to methadone maintenance treatment were excluded from the analysis of treatment success because they could arguably be perceived as having successful or unsuccessful courses of treatment; 358 patients remained. On admission, patients who were older, employed, and self-maintained with illicit buprenorphine had significantly higher odds of treatment success, whereas those of African American or Hispanic/Latino race had significantly lower odds of treatment success. Of patients remaining

Table 4. Urine Drug Tests During 12 Months[a] for Opioid-Dependent Patients Retained in Office-Based Opioid Treatment in Primary Care

Variable	No. (%) of Patients by Month			
	3	6	9	12
Opioids, total[b]				
Tested	263 (100)	220 (100)	189 (100)	169 (100)
Negative	249 (94.7)	207 (94.1)	176 (93.1)	161 (95.3)
Positive	14 (5.3)	13 (5.9)	13 (6.9)	8 (4.7)
Cocaine, total				
Tested	263 (100)	220 (100)	189 (100)	169 (100)
Negative	249 (94.7)	211 (95.9)	180 (95.2)	165 (97.6)
Positive	14 (5.3)	9 (4.1)	9 (4.8)	4 (2.4)

[a] Percentages are derived from how many patients were still enrolled at each time point who had at least 1 test result, using the last test before the time point.
[b] Oxycodone testing started on June 1, 2006.

in treatment at 12 months, 154 of 169 (91.1%) were no longer using illicit opioids or cocaine based on the results of urine drug testing (**Table 4**).

PROGRAM ACTIVITY

On average, the program coordinator received 20 calls per week requesting OBOT. An NCM saw 75 patients per week, and each OBOT physician prescribed to an average of 35 patients (range, 13-68) per month.

COMMENT

This large observational study of OBOT with buprenorphine in an academic medicine practice serves as a treatment model for facilitating access and improving outcomes in patients with opioid addiction. Collaborative care with NCMs resulted in feasible initiation and maintenance of buprenorphine for most patients (51.3%) admitted, which is comparable to previous studies in primary care settings involving smaller numbers of patients with 6- to 12-month follow-up.[9-15]

The collaborative care model ensured program compliance with federal laws[24] (eg, limits to the number of patients treated per OBOT physician) while maintaining access to a large number of patients. OBOT is ideal for collaborative care because much of the clinical work (eg, assessment, medication management, and monitoring) is based on established protocols.[23] Having relevant clinical data (eg, nursing assessments, documentation of treatment adherence, and urine drug test results) before physician visits allowed efficient use of physician time to focus on patient management (eg, dose adjustments and continued maintenance vs taper). In this particular setting, collaborative care allowed academic generalist physicians with research, administrative, and part-time clinical responsibilities to treat a large number of patients, many of whom had complex psychosocial needs. As described, the NCM was central to day-to-day clinical care with daily open access to address the myriad needs (eg, housing, employment, and health insurance) of this complicated population. This model satisfied a key principle of coordinated care by "assuring accessibility, continuity and high quality care that includes effective communication with, education of, and outreach to patients."[25(p404)]

Other important factors may have affected the outcomes of this intervention. Open communication between the NCM and addiction counselors improved patients' ability to comply with this essential element of good addiction care. Access to methadone maintenance allowed the program to safely transfer patients who required a more structured treatment modality.

We identified several preadmission factors associated with treatment success, some of which were consistent with the finding of previous studies and some that were not. Similar to previous reports, patients who were older[12] and employed[11,15] had more successful outcomes. The finding that self-maintaining with illicit buprenorphine on admission predicted success has not been previously published. These patients reported self-treating their opioid addiction with illicit buprenorphine rather than trying to get high. This suggests that these patients were highly motivated to obtain a dependable and health insurance–covered source of buprenorphine and to avoid relying on the illicit market. This finding is similar to that of other studies[26,27] that found that patients who self-treated with illicit methadone were more likely to have positive treatment outcomes. We did not find that cocaine use and psychiatric illness were associated with worse outcomes. This suggests that patients with these co-occurring issues should not be excluded from consideration for office-based treatment. We found that African American or Hispanic/Latino race lowered the odds of treatment success in our program. This finding requires confirmation and further study to better understand the ways that race and clinic structure may affect success in OBOT.

Several limitations of this study should be considered. Although data were collected prospectively using a medical record designed for OBOT patients, the study was retrospective, examining patients from a clinical program. In addition, follow-up information was not available after patient departure from the program. However, of patients who left the program earlier than 12 months, the reasons for leaving (ie, successful or unsuccessful treatment or methadone maintenance treatment transfer) were known. Urine drug test protocols also changed over time: testing for semisynthetic and synthetic opioids was not standardized un-

til June 1, 2006, so early urine results may underestimate prescription opioid abuse. However, once prescription opioid testing began, the rates of positive urine test results for illicit opioids did not change. This study did not have a control group, but this was acceptable because its purpose was not to retest the efficacy of buprenorphine treatment but to evaluate the feasibility of delivering this known effective treatment using a collaborative care model. Lastly, an experienced, skilled NCM played an essential role in caring for patients, and the ability to generalize such a model may depend on skills of such key individuals. However, the issue simply speaks to the need for the training of a nursing workforce with skills in caring for patients with addictions. In an effort to increase nursing involvement in buprenorphine treatment, a federally supported "Guide for Nurses" has recently been developed.[28]

In conclusion, OBOT can be offered effectively in a primary care practice using a collaborative care model. In this model, heavily reliant on NCMs, patient-level outcomes were comparable to those derived from other physician-centered approaches. This study of collaborative care adds to the growing body of evidence that office-based treatment of opioid addiction is feasible in primary care settings.

Accepted for Publication: August 19, 2010.
Correspondence: Daniel P. Alford, MD, MPH, Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center, 801 Massachusetts Ave, Second Floor, Boston, MA 02118 (dan.alford@bmc.org).
Author Contributions: *Study concept and design:* Alford, LaBelle, Kretsch, Bergeron, Botticelli, and Samet. *Acquisition of data:* Alford, LaBelle, Kretsch, and Bergeron. *Analysis and interpretation of data:* Alford, Kretsch, Winter, and Samet. *Drafting of the manuscript:* Alford, LaBelle, Kretsch, Bergeron, Winter, and Samet. *Critical revision of the manuscript for important intellectual content:* Alford, LaBelle, Kretsch, Bergeron, Winter, Botticelli, and Samet. *Statistical analysis:* Winter. *Obtained funding:* LaBelle and Botticelli. *Administrative, technical, and material support:* Alford, LaBelle, Kretsch, Bergeron, Botticelli, and Samet. *Study supervision:* Alford and Botticelli.
Financial Disclosure: None reported.
Funding/Support: Support for these programs and their evaluation was provided by the Massachusetts Department of Public Health's Bureau of Substance Abuse Services and HIV/AIDS Bureau. Dr Samet receives support from grant K24-AA015674 from the National Institute on Alcohol Abuse and Alcoholism.
Previous Presentations: Preliminary results of this study were presented at the 27th Annual Meeting of the Society of General Internal Medicine; May 15, 2004; Chicago, Illinois; and at the 28th Annual Meeting of the Association for Medical Education and Research in Substance Abuse; November 11, 2004; Washington, DC.

REFERENCES

1. McLellan AT, Lewis DC, O'Brien CP, Kleber HD. Drug dependence, a chronic medical illness: implications for treatment, insurance, and outcomes evaluation. *JAMA*. 2000;284(13):1689-1695.
2. Leshner AI. Science-based views of drug addiction and its treatment. *JAMA*. 1999; 282(14):1314-1316.
3. Substance Abuse and Mental Health Services Administration. *Results From the 2006 National Survey on Drug Use and Health: National Findings*. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2007. Office of Applied Studies, NSDUH series H-32, DHHS publication SMA 07-4293.
4. Gibson A, Degenhardt L, Mattick RP, Ali R, White J, O'Brien S. Exposure to opioid maintenance treatment reduces long-term mortality. *Addiction*. 2008;103 (3):462-468.
5. Mattick RP, Breen C, Kimber J, Davoli M. Methadone maintenance therapy versus no opioid replacement therapy for opioid dependence. *Cochrane Database Syst Rev*. 2009;(3):CD002209.
6. National Consensus Development Panel on Effective Medical Treatment of Opiate Addiction. Effective medical treatment of opiate addiction. *JAMA*. 1998; 280(22):1936-1943.
7. Fiellin DA, O'Connor PG. Clinical practice: office-based treatment of opioid-dependent patients. *N Engl J Med*. 2002;347(11):817-823.
8. Alford DP, LaBelle CT, Richardson JM, et al. Treating homeless opioid dependent patients with buprenorphine in an office-based setting. *J Gen Intern Med*. 2007;22(2):171-176.
9. Colameco S, Armando J, Trotz C. Opiate dependence treatment with buprenorphine: one year's experience in a family practice residency setting. *J Addict Dis*. 2005;24(2):25-32.
10. Fiellin DA, Moore BA, Sullivan LE, et al. Long-term treatment with buprenorphine/naloxone in primary care: results at 2-5 years. *Am J Addict*. 2008;17(2):116-120.
11. Magura S, Lee SJ, Salsitz EA, et al. Outcomes of buprenorphine maintenance in office-based practice. *J Addict Dis*. 2007;26(2):13-23.
12. Mintzer IL, Eisenberg M, Terra M, MacVane C, Himmelstein DU, Woolhandler S. Treating opioid addiction with buprenorphine-naloxone in community-based primary care settings. *Ann Fam Med*. 2007;5(2):146-150.
13. Parran TV, Adelman CA, Merkin B, et al. Long-term outcomes of office-based buprenorphine/naloxone maintenance therapy. *Drug Alcohol Depend*. 2010; 106(1):56-60.
14. Soeffing JM, Martin LD, Fingerhood MI, Jasinski DR, Rastegar DA. Buprenorphine maintenance treatment in a primary care setting: outcomes at 1 year. *J Subst Abuse Treat*. 2009;37(4):426-430.
15. Stein MD, Cioe P, Friedmann PD. Buprenorphine retention in primary care. *J Gen Intern Med*. 2005;20(11):1038-1041.
16. Netherland J, Botsko M, Egan JE, et al; BHIVES Collaborative. Factors affecting willingness to provide buprenorphine treatment. *J Subst Abuse Treat*. 2009; 36(3):244-251.
17. Walley AY, Alperen JK, Cheng DM, et al. Office-based management of opioid dependence with buprenorphine: clinical practices and barriers. *J Gen Intern Med*. 2008;23(9):1393-1398.
18. Barry DT, Irwin KS, Jones ES, et al. Integrating buprenorphine treatment into office-based practice: a qualitative study. *J Gen Intern Med*. 2009;24(2):218-225.
19. Hunt JS, Siemienczuk J, Pape G, et al. A randomized controlled trial of team-based care: impact of physician-pharmacist collaboration on uncontrolled hypertension. *J Gen Intern Med*. 2008;23(12):1966-1972.
20. DiPiero A, Dorr DA, Kelso C, Bowen JL. Integrating systematic chronic care for diabetes into an academic general internal medicine resident-faculty practice. *J Gen Intern Med*. 2008;23(11):1749-1756.
21. Saitz R, Larson MJ, Labelle C, Richardson J, Samet JH. The case for chronic disease management for addiction. *J Addict Med*. 2008;2(2):55-65.
22. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed, text revision. Washington, DC: American Psychiatric Association; 2000.
23. Center for Substance Abuse Treatment. *TIP 40: Clinical Guidelines for the Use of Buprenorphine in the Treatment of Opioid Addiction*. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2004. DHHS publication (SMA) 07-3939.
24. Drug Addiction Treatment Act of 2000. Pub L 106-310, 114 Stat 1223-1227 (2000).
25. Blue Ribbon Panel of the Society of General Internal Medicine. Redesigning the practice model for general internal medicine: a proposal for coordinated care: a policy monograph of the Society of General Internal Medicine. *J Gen Intern Med*. 2007;22(3):400-409.
26. Sohler NL, Li X, Kunins HV, et al. Home- versus office-based buprenorphine inductions for opioid-dependent patients. *J Subst Abuse Treat*. 2010;38(2):153-159.
27. Cunningham C, Giovanniello A, Sacajiu G, et al. Buprenorphine treatment in an urban community health center: what to expect. *Fam Med*. 2008;40(7):500-506.
28. Center for Substance Abuse Treatment. *TAP 30: Buprenorphine: A Guide for Nurses*. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2009. DHHS publication (SMA) 09-4376.

Original Article

Assessing the Impact of Patient Navigation

Prevention and Early Detection Metrics*

Tracy A. Battaglia, MD, MPH[1]; Linda Burhansstipanov, MSPH, DrPH[2]; Samantha S. Murrell, MPH[1]; Andrea J. Dwyer, BS[3]; and Sarah E. Caron, MPH[1]; on behalf of The Prevention and Early Detection Workgroup from the National Patient Navigation Leadership Summit

BACKGROUND. The lack of comparable metrics to evaluate prevention and early detection patient navigation programs impeded the ability to identify best practices. **METHODS.** The Prevention and Early Detection Workgroup of the Patient Navigation Leadership Summit was charged with making recommendations for common clinical metrics specific to the prevention and early detection phase of the cancer care continuum. The workgroup began with a review of existing literature to characterize variability in published navigation metrics; then developed a list of priority recommendations that would be applicable to the range of navigation settings (clinical, academic, or community-based). **RESULTS.** Recommendations for researchers and program evaluators included the following: 1) Clearly document key program characteristics; 2) Use a set of core data elements to form the basis of your reported metrics; and 3) Prioritize data collection using methods with the least amount of bias. **CONCLUSIONS.** If navigation programs explicitly state the context of their evaluation and choose from among the common set of data elements, meaningful comparisons among existing programs should be feasible. ***Cancer*** **2011;117(15 suppl):3553–64.** © *2011 American Cancer Society.*

KEYWORDS: patient navigation, prevention and control, early detection of cancer, cancer screening, program evaluation, measures.

INTRODUCTION

Cancer control begins with primary and secondary prevention efforts which aim to reduce cancer incidence and advanced disease, respectively. The evidence is clear that certain cancers—those caused by tobacco use, viruses, or sun exposure, for example—can be prevented completely. Regular use of proven screening modalities, such as Pap tests for cervical cancer and colonoscopy for colorectal cancer, also result in prevention through the removal of precancerous lesions. Other screening tests can detect cancers of the breast, colon, rectum, cervix, prostate, oral cavity, and skin at early stages and translate into a direct mortality benefit when abnormal screening is followed by prompt diagnosis and treatment. Mounting evidence suggests that the delivery of prevention and early detection (PED) services are responsible for a substantial portion of the documented reduction in both cancer incidence[1] and mortality[1,2] in the United States.

It is also well documented that not all populations benefit equally from these prevention efforts, in part because our current healthcare delivery system does not provide consistent, high-quality care to all.[3] Whether defined by age, gender, race, insurance status, geographic location, or comorbid medical condition, certain populations face significant barriers to accessing timely and quality cancer PED services consistently, if at all.[4-6] Patient Navigation, which targets barriers faced by

Corresponding author: Tracy A. Battaglia, MD, MPH, 801 Massachusetts Avenue, Crosstown Building, Suite 470, Boston, MA 02118; Fax (617) 638-8096; tracy.battaglia@bmc.org

[1]Women's Health Unit, Section of General Internal Medicine, Department of Medicine, and Women's Health Interdisciplinary Research Center, Boston University School of Medicine, Boston, Massachusetts; [2]Native American Cancer Research Corporation, Pine, Colorado; [3]Colorado School of Public Health, University of Colorado Cancer Center, Denver, Colorado

This conference and supplement were cosponsored by Pfizer Oncology, Livestrong (Lance Armstrong Foundation), Susan G. Komen for the Cure, the Oncology Nursing Society (ONS), the American College of Surgeons Commission on Cancer, the American Cancer Society, and AstraZeneca.

The articles in this supplement are based on presentations at the "National Patient Navigator Leadership Summit"; March 23-24, 2010; Atlanta, GA.

The opinions or views expressed in this supplement are those of the authors, and do not necessarily reflect the opinions or recommendations of the publisher, the editors, the University of Illinois at Chicago, the American Cancer Society, or the Ralph Lauren Center for Cancer Care and Prevention.

**National Patient Navigator Leadership Summit (NPNLS): Measuring the Impact and Potential of Patient Navigation, Supplement to Cancer.*

DOI: 10.1002/cncr.26267, **Received:** February 24, 2011; **Revised:** May 2, 2011; **Accepted:** May 6, 2011, **Published online** July 20, 2011 in Wiley Online Library (wileyonlinelibrary.com)

vulnerable populations in accessing timely, quality cancer care,[7] was designed to address the critical disconnect between the discovery and delivery of life-saving cancer care services. In fact, the first patient navigation program was started in Harlem, New York, to increase the delivery of mammography screening to Black women who were too often presenting with advanced cancer as a result of a lack of screening.[8] This groundbreaking work used lay navigators from the local community to help at-risk women overcome barriers to accessing screening and diagnostic services and resulted in profound improvements in breast cancer care.[8] Since then, a growing number of studies documenting the promise of navigation have resulted in its widespread adoption as a means to deliver PED services.[9-15]

As navigation becomes integrated into standard cancer care services across the country,[16] the lack of comparable metrics to evaluate these programs in different settings with diverse target populations impedes our ability to identify best practices and realize the full potential of this promising intervention. Thus, we aim here to provide recommendations for researchers and program evaluators to consider adopting when measuring the impact of their PED navigation programs. The intent is to facilitate consistent use of priority metrics, including process and intermediate outcome measures, that document the type and quality of work performed by prevention and early detection Patient Navigators (PN) working in diverse settings (clinical, academic, community-based organizations). Through the use of such measures, public health and health reform policies may be generated to provide reimbursement for services that ensure the delivery of timely, quality cancer prevention.

METHODS

In March 2010, the American Cancer Society hosted the first National Patient Navigation Leadership Summit, where it convened cancer clinicians, researchers, and practicing public health experts to develop a national evaluation agenda for patient navigation. The Prevention and Early Detection (PED) Workgroup was charged with making recommendations for common clinical metrics specific to the prevention and early detection phase of the cancer care continuum. The workgroup comprised 10 individuals, representing community-based organizations, clinical programs and academia with decades of experience implementing and evaluating patient navigation programs across diverse populations. The workgroup began with a review of existing literature to characterize variability in published metrics, then through discussion and consensus developed a list of priority recommendations.

In early 2010, the Summit Planning Committee conducted a comprehensive review of the navigation literature to guide discussion at the March meeting. The PED work group updated the literature review in October of 2010. We searched the Pubmed database to identify original articles published any year, in English, using the key terms "patient navigation," "patient navigator," "navigation," "navigator," or "case management." We also searched the references of each publication for additional relevant literature. In keeping with the scope of navigation as outlined by Dr. Freeman at the Summit, we included only intervention studies where navigators actively link patients to clinical services. Educational or outreach navigation for the delivery of prevention education in community settings were excluded. Likewise, studies that used labels for functional navigators, such as community outreach worker, community health advisor/aide, promotores, or lay health advisor/educator, may have been excluded. We present here findings from the synthesis of 32 published articles we believed exemplified the breadth of published metrics. Although not meant to be an exhaustive review of the extensive literature, the studies included are representative of the variability in existing metrics.

FINDINGS FROM LITERATURE REVIEW

Most studies target breast and colorectal cancer, with fewer targeting cervical, lung and/or prostate cancer. Reported clinical outcome metrics fall into 2 discrete points along the continuum of cancer care: 1) screening and 2) diagnosis, while the remaining metrics focus on the processes specific to the navigation program. To date, no patient navigation intervention study has reported final endpoints such as survival or mortality. Rather, the current literature focuses mainly on intermediate clinical outcomes in the form of the delivery of recommended cancer prevention services. Only 2 studies[17,18] document a potential mortality benefit in the form of a stage shift at the time of diagnosis. As discussed below and summarized in Tables 1 and 2, there exists wide variation in both the reporting of nonmodifiable program characteristics as well as how study outcome metrics are defined and reported.

Screening

We reviewed 20 navigation studies that targeted cancer screening as an outcome (Table 1). We include studies

Table 1. Variability in Published Patient Navigation Studies: Screening Metrics

Study	Setting		Disease	Eligibility Criteria	Navigation Mode	Outcome Metric	Follow-Up Period	Data Collection
Weber, 1997	Urban	Clinical	Breast	Age 52 - 77 Last mammo > 24 mo	Telephone	Receipt of mammo	10 mo	Medical record review
Burhansstipanov L, 2000	Urban	Community	Breast	Age ≥39 Last mammo > 18 mo	Face-to-face Telephone	Receipt of screening test	NR	Self-report
Dignan, 2005	Urban	Community	Breast	Age 40+ Last mammo > 18 mo	Face-to-face Telephone	Receipt of mammo	12 mo	Self-report
Jandorf, 2005	Urban	Clinical	Colorectal	Age 50+ FOBT > 1 yr FS or BE >5 yr Colonoscopy >10 yr	Telephone	Receipt of CRC screening	6 mo	Medical record review
Dietrich, 2006	Urban	Clinical	Breast Cervical Colorectal	Age 50 - 69 Overdue screening	Telephone	Adherence to recommended screening	18 mo	Medical record review
Ford, 2006	Urban	Clinical	Prostate Lung Colorectal	Age ≥55	Telephone	Receipt of next scheduled screening test	Time to next trial screening	Medical record review
Nash, 2006	Urban	Clinical	Colorectal	All colonoscopies	NR	Receipt of colonoscopy	11 mo	Medical record review
Paskett, 2006	Rural	Community	Breast	Age ≥ 41 Last mammo > 12 mo	Face-to-face Telephone	Receipt of mammo Barriers to mammo	12 mo	Medical record review
Myers, 2008	NR	Clinical	Colorectal	Age 50 - 79 Last visit < 24 mo	Telephone	Receipt of CRC screening	6 mo	Self-report Medical record review
Percac-Lima, 2008	Urban	Clinical	Colorectal	Age 52 - 79 FOBT > 1 yr FS or BE >5 yr Colonoscopy >10 yr	Telephone	Receipt of CRC screening Barriers to CRC screening	9 mo	Medical record review
Clark, 2009	Urban	Clinical	Breast	Age 18 - 75	Telephone	Receipt of mammo Maintenance screening behavior Timely adherence to diagnostic resolution	3 yr	Medical record review
Fernandez, 2009	Rural	Community	Breast Cervical	Age ≥50 Farm worker status	Face-to-face	Receipt of mammo, pap test	6 mo	Self-report Medical record review for validity
Han, 2009	NR	Community	Breast	Age 40+ Last mammo > 24 mo	Face-to-face Telephone	Receipt of mammo	6 mo	Self-report
Lasser, 2009	Urban	Clinical	Colorectal	Age 52 - 80 No CRC screening	Telephone	Receipt of CRC screening # patients contacted # hours spent navigating	6 mo	Medical record review
Ma, 2009	NR	Community	Colorectal	Age 50+ ACS guidelines for CRC screening	Face-to-face	Receipt of CRC screening Barriers to CRC screening	12 mo	Self-report verified with physician's office
Burhansstipanov L, 2010	Urban	Community	Breast	Age ≥ 39 Last mammo > 18 mo	Face-to-face Telephone	Receipt of screening mammo	NR	Self-report
Phillips, 2010	Urban	Clinical	Breast	Age 51 - 70 Last mammo > 18 mo	Telephone	Adherence to recommended screening	24 mo	Medical record review
Wang, 2010	Urban	Community	Cervical	Age 18+ Last pap > 12 mo	Face-to-face	Receipt of pap test	12 mo	Self-report

ACS, American Cancer Society; CRC, colorectal cancer; NR, not reported.

Table 2. Variability in Published Patient Navigation Studies: Diagnostic Metrics

Study	Setting		Disease	Eligibility Criteria	Navigation Mode	Outcome Metric	Follow-Up Period	Data Collection
Ell, *Cancer Pract,* 2002	Urban	Clinical	Breast	Age: None Prescribed follow-up screening/ diagnostic test	Telephone	Receipt of diagnostic resolution Timely adherence to diagnostic resolution Barriers to diagnostic resolution	NR ACR 4&5: 60 days ACR 3: 240 days	Appointment records Self-report
Ell, *J Wmn's Hlth Gend Based Med* 2002	Urban	Clinical	Cervical	Age: None LGSIL & HGSIL Prescribed follow-up screening/ diagnostic	Telephone	Receipt of diagnostic resolution Timely adherence to diagnostic resolution	12 mo 30 days	Medical record review
Oluwole, 2003	Urban	Clinical	Breast	Age: None Clinic patients	NR	Stage at diagnosis	Retrospective, cross-sectional	Medical record review
Battaglia, 2007	Urban	Clinical	Breast	Age: >18 Referred for evaluation	Telephone	Timely adherence to diagnostic resolution	120 days	Medical record review Self-report
Ell, 2007	Urban	Hospital	Breast	Age: None ACR 3-5	Telephone	Receipt of diagnostic resolution Timely adherence to diagnostic resolution Barriers to diagnostic resolution	8 mo ACR 4&5: 60 days ACR 3: 240 days	Medical record review
Ferrante, 2007	Urban	Clinical	Breast	Age: $\geq$ 21 BIRADS 4 & 5	Telephone Face-to-face	Timely adherence to diagnostic resolution Time to diagnostic resolution	60 days Unclear	Medical record review
Gabram, 2008	Urban	Clinical	Breast	Age: None Clinic patients	NR	Stage at diagnosis	Retrospective, cross-sectional	Medical record review
Palmieri, 2009	Urban	Clinical	Breast	Age: None 200% FPL	NR	Timely adherence to diagnostic resolution Receipt of diagnostic resolution	60 days NR	Medical record review
Bastani, 2010	Urban	Clinical	Breast	Age: None Referred to surgery or radiology for breast abnormality	Telephone	Timely adherence to diagnostic resolution	6 mo	Medical record review

ACR, American College of Radiology; BIRADS, Breast Imaging Reporting and Data System; FPL, Federal Poverty Level; NR, not reported.

with community- and clinically based navigators in urban[19-25] and rural settings.[14,26] The studies targeted diverse populations, including American Indians,[14,15,19,26] Korean-Americans,[27] Chinese women,[28] Latinas,[14,29] Blacks,[14,26,30,31] non-English speaking,[32] poor Whites,[14] low-income.[32-34] Few programs were comprehensive targeting multiple cancer sites,[29,30,33] while most target only 1 disease-specific screening.[14,15,19-28,31,32,34-36] Even among studies targeting the same disease, eligibility criteria for inclusion in programs vary, including the age of participants and the time since their last screening. For example, 1 mammography screening navigation study included women 52 to 77 years who had not had a mammogram in the previous 2 years,[21] while another included women over 40 years whose last mammogram was 12 or more months prior.[26]

Most studies document receipt of a screening test as the goal of navigation and report the outcome simply as a screening rate, defined as the proportion of eligible subjects who complete a recommended test, such as a mammogram, Pap test, or colonoscopy, during the intervention period. The range of the intervention period across studies was wide, such that the time subjects were followed to assess the outcome varied from 6 months[21,24,32-34] to 3 years.[30] The most common follow-up period was 6 months.[22,23,27,29,36] Two studies document adherence to recommended screening[20,33] as the goal of navigation and report the outcome as an adherence rate, defined as the proportion of eligible subjects who are up-to-date with a screening test as recommended by an existing guideline or standard. These 2 breast navigation studies differed in how they defined "adherent"; 1 used United States Preventive Task Force (USPTF) guidelines,[33] another Healthcare Effectiveness Data and Information Set (HEDIS) criteria.[20] Only 1 study reports maintenance screening behavior,[31] which was defined as the percentage of annual mammograms that were actually obtained during the study period. Data collection methods were either self-reported behaviors[13-15,19,27-29,35] or objective evidence from medical record review.[20-24,26,31-33,36]

Diagnosis

Of the 13 studies included targeting the diagnostic phase of the cancer care continuum, 12 targeted breast cancer diagnosis[8-10,12,13,17,18,31,37-40] while only 1 targeted cervical cancer.[11] As shown in Table 2, we include studies with a range of program settings that target diverse populations with variable eligibility criteria. Similar to screening navigation studies, the range of the intervention period across studies was wide, and the time subjects were followed to assess the outcome varied. Studies report 4 clinical metrics at the point of diagnostic evaluation: 1) receipt of diagnostic resolution,[8-11,38,40] 2) time to diagnostic resolution,[8,37-39] 3) timely adherence to diagnostic resolution,[9-13,31,37,38] and less commonly 4) stage at diagnosis.[17,18]

Five studies report receipt of diagnostic resolution[8-11,38,40] as the goal of navigation. These studies present this outcome simply as a resolution rate, defined as the proportion of eligible subjects who complete diagnostic testing during the intervention period. The majority of studies reviewed report timeliness of diagnostic care as the goal of patient navigation. These studies report timely in 2 distinct ways: either 1) the time to diagnostic resolution[8,37-39] as a continuous outcome, or 2) the rate of timely adherence to diagnostic resolution as a dichotomous outcome.[9-13,31,37,38]

The most striking finding in reviewing these metrics is the lack of a consistent definition for what constitutes "diagnostic resolution" or the "timely" diagnostic interval. Most studies use the date the abnormal screening test was performed as the index event or start date.[9,10,31,37,38] However, there is widespread variability in the data point indicating diagnosis, diagnostic resolution or adherence to recommended follow-up, ranging from the date of arrival to first diagnostic clinical visit[12] to the actual date a tissue sample was obtained.[8] When tissue diagnosis is not recommended, studies vary in reporting how a "diagnostic resolution" is determined. For example, 1 study reports the endpoint as "until negative mammogram, benign biopsy, 6 month follow-up test, or start of cancer treatment",[10] while other studies only include benign or malignant tissue as a diagnosis.[39] There is similar variability in how investigators define "timely" ranging from 60 to 180 days.[9,13]

The Patient Navigation Research Program, a collaborative multisite research program designed to evaluate the efficacy of navigation after abnormal cancer screening, developed a set of "common" data points using the National Comprehensive Cancer Network (NCCN) guidelines as the major focus of clinical outcomes.[41] While the results of this program are not yet published, the PNRP is the largest study to date on PED navigation. In their program, diagnostic resolution is defined as completion of the diagnostic test that results in a diagnosis or clinical evaluation that determines that no further evaluation is indicated. For example, a colonoscopy with biopsy confirming a malignant polyp or a colonoscopy in which no malignant lesion is identified would both serve as a diagnostic resolution.

The 2 studies reporting breast cancer stage at diagnosis as the outcome similarly reported population level data to assess the impact of a navigation program targeting individuals.[17,18] These studies suggest a positive impact of patient navigation; however, due to the methods used, a causative association cannot be determined.

Process Metrics

In addition to intermediate clinical outcomes, 6 studies included here report metrics that evaluate whether the intervention was implemented as intended. Five studies report navigator-documented barriers to care.[9,10,25,31,35] One study by Lasser et al documents the median number of contacts per patient and mean hours of telephone outreach per patient.[23] A descriptive study by Lin et al documented the types of barriers to care and time spent by the navigator.[42] The PNRP is collecting the following process metrics in their multi-site program: barriers to care identified by navigator, actions taken by the navigator, and details of navigation encounters such as type of encounter and time spent.[41] Only 1 of these studies[33] have examined the association between these process measures and their outcomes, which represents an area in critical need for further study.

Recommendations for PED Metrics

In keeping with the goal of having a common set of priority metrics for navigation programs to measure impact on individuals and populations, it would be ideal to have consistent study characteristics, including eligibility criteria, follow-up time intervals and outcome metrics. Obviously, this is not possible for several reasons. First, there are certain program characteristics that are inherently nonmodifiable such as program setting and the populations they serve. In addition, the specific needs of populations appropriately dictate the intended outcomes of navigation, the ideal mode of navigator contact or specific navigator activities. Finally, there is wide variability in the amount and type of resources available for evaluation efforts. Community programs wishing to conduct an evaluation of their program may well have fewer resources than a federally funded research project such as the PNRP. Regardless, the existing literature illuminates the need for consistency in reporting both modifiable and nonmodifiable program characteristics. Stating these clearly will facilitate meaningful program comparisons even in the absence of common outcome metrics.

Table 3. Reportable Program Characteristics for Prevention and Early Detection Navigation Programs

Construct	Common Metric
Setting	Urban v. Rural v. Suburban
	Clinical v. community
Eligibility criteria of patients	Age
	Race
	Primary language spoken
	Time since last screening
Mode of navigation	In person v. telephone
	# encounters per patient
	Time spent per patient
	# patients navigated
Follow-up interval	# months or years

Therefore, our first recommendation is to clearly document the following minimal set of program characteristics (Table 3):

1. *Program setting*. At a minimum knowledge of geographic settings such as urban, rural, suburban is an important distinction. More importantly, the system setting is essential to know when considering replicating a program. Beyond describing whether a program is community versus clinically-based, some detail on the specific area within a clinical setting (e.g. primary care versus radiology), or community setting (e.g. church versus YWCA) is important.
2. *Eligibility criteria of navigated subjects*. These programmatic elements are necessary to order to interpret the outcomes and their potential impact for other populations. Most important are age, race/ethnicity, primary spoken language and time since last screening.
3. *Mode of navigation*. The primary mode of delivering the navigation program is a minimal program element essential to comparing study findings. Specifically, did the navigator interact with their target community in person (in a community setting, in a clinical setting) or on the telephone? How many encounters did the navigator have over the course of the intervention? In addition, the amount of time spent per patient and overall navigator case load is important to document (with how many patients per week does the navigator interact, how many navigators worked with the same patient to access follow-up diagnostic services).
4. *Time interval of the follow-up period at which outcomes are assessed*. This detail is critical to

Table 4. Recommended Common Data Elements for Screening Metrics

Construct	Common Data Elements	Common Outcome Metrics
Receipt of screening test	A. Date enrolled into navigation B. Date referred for screening C. Date test scheduled (#1, #2, #3, etc) D. Date test completed E. Date test results are read/reported F. Date patient informed of test result	Completion of screening test *(Yes/No)* Timely completion of screening *(Yes/No)* • Must define "timely" Time to complete screening *(# days A - D)*
Adherence to single recommended screening interval	A. Name of professional guidelines that defines recommended screening (ie, USPSTF, NCCN) B. Date current test completed C. Date most recent screening test completed	Adherent to single recommended screening *(Yes/No)*
Adherence to longitudinal recommended screening = maintenance	A. Name of professional guidelines that defines recommended screening maintenance (ie, USPSTF, NCCN) B. Date current screening test completed C. Date most recent screening test completed D. Date past screening tests completed	Adherent to longitudinal screening *(Yes/No)* • Must define "longitudinal screening"

interpreting the meaning of a defined outcome. For example, it would important to know that 2 programs reporting a similar clinical outcome (ie, 90% of program participants completed their mammogram) each measured their outcome at different time intervals (ie, 1 year vs 6 months).

Many of the observed differences in published PED outcomes are not in the data elements collected, rather in either the nomenclature used to describe them or the analyses used to report them. Thus, defining a common set of *data elements*, rather than firm outcome metrics is a much more realistic approach and comprises our second set of recommendations. Prioritizing the collection of these data elements will allow for the variability inherent in navigation programs that target different communities and systems of care while also allowing for meaningful comparisons. From these data elements, common metrics to represent prevention and early detection constructs can be created (see Tables 4, 5, and 6).

Table 4 displays recommendations for common data elements that may be used to create a set of intermediate outcome metrics that fit within the constructs of screening. The first construct is *receipt of the screening test*. Documenting the core data elements to measure this construct allows for the reporting of dichotomous outcomes metrics like "completion of screening test (yes/no)" or "timely completion of screening test (yes/no)" as well as the continuous outcome of "time to complete screening". Either of these metrics may be reported using any or all of the common data elements outlined in Table 4. This measure is limited when comparing programs with different eligibility criteria and follow-up time periods. Thus, a more comparable construct to consider is *adherence to recommended screening*, which requires of course that a screening guideline (such as the USPSTF) be stated explicitly. This common metric allows for programs to compare their adherence rates across different populations. Because the full benefit of screening on survival is dependent on the longitudinal use of "routine" screening tests, or maintenance of screening over time, there should be an emphasis toward documenting screening maintenance behavior.

Table 5 displays recommendations for common data elements that may be used to create a set of intermediate outcome metrics that fit within the constructs of diagnostic outcomes. Common metrics for reporting the construct of *diagnostic resolution* must begin with a clear definition of which core data elements constitute diagnosis and/or resolution of the screening abnormality. Once this is clear, reporting of the dichotomous outcome metrics may include "completion of diagnostic resolution (yes/no)," "timely completion of diagnostic resolution (yes/no)" and the continuous outcome of "time to complete diagnostic resolution." These metrics may be reported using any or all of the common data elements outlined in Table 5. Resources and program intent will create variability in which data elements programs are interested in and capable of collecting. The priority should be to have an explicit and consistent definition of diagnostic *resolution* and to collect the date corresponding with that definition, as recommended by the PNRP.[41] When the diagnostic resolution is a diagnosis of cancer,

Table 5. Recommended Common Data Elements for Diagnostic Metrics

Construct	Common Data Elements	Common Outcome Metrics
Receipt of diagnosis or resolution of screening abnormality	A. Date index screening test performed B. Date patient informed of test result C. Date enrolled into navigation clinical evaluation D. Date of first scheduled diagnostic test/clinic visit • Date of second scheduled, third scheduled, etc E. Date of completion of first diagnostic test/clinic visit F. Date of completion of final diagnostic test/clinic visit G. Result of final test performed (cancer Yes/No) H. Date diagnostic test read/reported I. Date patient informed of test results	Completion of diagnostic resolution *(Yes/No)* • Must define "diagnosis" • Must define "resolution" Time to completion of diagnostic resolution[42] *(# days A - F)* Timely completion of diagnostic resolution *(Yes/No)* • Must define "timely"
Adherence to recommended diagnostic testing	A. Name of professional organization/guidelines B. Type of test resulting in diagnostic resolution	Adherent to recommended diagnostic testing *(Yes/No)* Completion of appropriate test *(Yes/No)* • Must define appropriate test (eg, percutaneous biopsy v. open biopsy)
Stage at diagnosis	A. TNM Classification of Malignant Tumors cancer staging criteria[56]	Stage 0-4

Table 6. Recommended Common Data Elements for Process Metrics

Construct	Common Data Elements	Common Outcome Metrics
Phase of cancer care targeted by navigation program	Outreach / Screening or Diagnostic clinical visit / Follow-up	Phase of cancer care
Adherence to scheduled clinical visit	A. Date of appointment B. Type of appointment C. Status of appointment • Arrive • No show • Cancel • Reschedule	Adherent to appointment *(Yes/No)*
Caseload	A. # of patients navigated per navigator B. Time spent per patient (minutes, hours) C. # days in navigation	Navigator Caseload *(# patients / time period)*
Communication	A. Encounter type: in person, phone, letter B. Interpreter used *(Yes/No)* C. Date of first encounter D. Date of last encounter	Mode of communication
Barriers	A. See PNRP Methods paper[41]	# of barriers / per patient Type of barriers
Actions	A. See PNRP Methods paper[41]	# of actions / per barrier or per patient Type of actions

metrics such as stage at diagnosis are also important to record. Another recommended construct that is often omitted from program evaluation is *adherence to recommended testing*, as determined by documentation of the type of diagnostic test performed in approaching diagnostic resolution. This common metric, *completion of appropriate test*, is another measure of quality to ensure populations have access to appropriate diagnostic testing.

Our third set of recommendations call for a minimal set of process data elements (Table 6). Process measures are intended to measure whether navigation was delivered as planned or designed. Without these details, replication of programs with successful outcomes is not possible.

Figure 1. Native American Cancer Research Corporation tool for documenting common process data elements: Navigator Actions.

Knowledge of the specific components of a navigation program is necessary to apply lessons learned from 1 program to the next. The PED Working Group identified 3 distinct phases of PED where processes of navigation may differ: 1) outreach/promotion (helping community understand the need and availability of cancer screening); 2) support during clinical appointments and tests; and 3) tracking and follow-up after appointments/tests completed.

At a minimum, programs should document which phase(s) of PED their navigators address, as this captures broadly the types of activities involved in the navigation program.

In addition, we strongly recommend that programs document clinical appointment data to report health services process measures related to adherence to scheduled clinical visits. Also of importance to program function and impact is the number of patients navigated (over some specified time period: daily, weekly, or monthly) and the time spent with individual patients. From this information measures of caseload may be created. Mode of communication and whether an interpreter was used in an encounter is another important process measure. Documenting the date of last navigator encounter ensures a way to attribute the screening outcome to navigation. For example, we would not want to attribute a screening outcome to navigation if there has been no contact with a navigator in the prior 12 months.

Considering that addressing barriers to care are at the very center of the conceptual model of navigation, it is essential to measure these barriers along with navigator activities, or actions, taken to address them. Creating an optimal set of patient-level barriers to care is challenging given the specific needs of diverse populations, as barriers in 1 community may be vastly different from barriers in another. Freund et al describe recommendations for barriers used by the PNRP[41] and provide a framework for documenting navigation activities that would facilitate meaningful comparisons. The Native American Cancer Research Corporation (NACR) provides another example of documenting barriers and actions routinely used in their program (Fig. 1).[15] Finally, documenting healthcare usage along the screening process is an alternative way of capturing benefits of navigation, such as reduction in rates of missed appointments.

Our fourth and final set of recommendations is related to data collection efforts. Data elements may be collected using patient self-report, navigator logs, clinical

data sources, and/or objective observation. At a minimum, programs should document their data source, given the limitations/strengths of these various sources. In a research context, it would be inappropriate to have navigators administer outcome assessments for their own patients as it would introduce potential bias. Whereas it would be acceptable to have navigators document process measures, programs should avoid using navigators to document clinical outcomes without extensive quality assurance in place.

PN daily logs are an obvious source for process measures. Electronic programs can be used for those PN who have access to computers and/or the Internet. Tremendous effort should be made to ensure the layout consists of closed question format or checkboxes that address most prevalent responses with an "other" category that allows for text input. These lists or checkboxes should include space to document the amount of time the PN spent doing each task, or better yet checkboxes with time intervals.

Patient self-reported screening behaviors are often inaccurate[43] and how the questions are asked may influence the responses. However, if patient self report is used, phrasing of questions should be drawn from standardized, validated instruments such as the National Health Interview Survey (NHIS), the Behavioral Risk Factor Surveillance System (BRFSS), or the National Medical Care Expenditure Survey (NMEPS). Likewise, there are differences in the types of responses when such instruments are administered face-to-face, over the phone, completed by the patient, use of CADI (Computer Aided Design Instrument) systems and/or through the Internet.[44]

While objective observation methods of patients and navigators have been developed,[45] most programs will not have the resources to use them. With federal mandates requiring transition to electronic medical records, there is tremendous opportunity to use objective clinical data sources to measure these outcomes and should be the standard for navigation programs to aspire. For example, electronic medical records may be queried for the presence of screening reports or as a means to complete certain data points, while electronic registration systems may be queried to report adherence outcomes for scheduled appointments.

DISCUSSION

Patient navigation programs that target the prevention and early detection spectrum of care share similar goals, yet vary widely in how they document their success. Differences in program structure, population needs, outcomes of interest, and reported evaluation metrics make cross-study comparisons impossible. However, a review of the literature suggests that a common set of evaluation metrics relevant to multiple stakeholders can be developed. Based on a synthesis of existing navigation literature and expert consensus, we present here a set of 4 recommendations related to measuring and reporting PED navigation program success so that dissemination of the evidence may be used to delineate best practices in the design of care processes across diverse settings.

Our recommendations call for a core set of quality indicators that measure the intent of navigation—to bridge the critical disconnect between the discovery and delivery of life saving cancer care services. Knowledge of basic program characteristics is the starting point to contextualize comparisons between programs. While clinical outcome measures of quality (eg, stage of diagnosis or mortality) are generally more difficult or not feasible to measure, we provide a framework of common data elements that may be used to report a common set of intermediate clinical metrics. Equally important, we provide recommendations for collecting and reporting process measures (activities performed while receiving care) which are the most frequently used quality indicators,[47,48] because they are sensitive, unambiguous, and easily measured.[49-51] Our review of the literature highlights the lack of evidence linking these processes to clinical outcomes, making these data elements of high priority for future study as process measures should be associated to outcome measures for effective quality assessment.[52,53]

A limitation of this study was focusing the literature review to PubMed, certainly the most commonly used database. However, some navigation projects and studies using labels for staff who function as navigators linking individuals from community to the health system, such as community outreach worker, community health advisor/aide, promotores, or lay health advisor/educator, may have been excluded. Likewise, several of the latter articles are accessible through publication databases that focus on education, social work, and psychology that may or may not also be within PubMed. As a result, this article is less inclusive of community-based and academic-based navigation programs and emphasizes the clinical settings. Regardless, most of the recommended measures presented here are relevant to navigators working in these other settings as well.

Priorities should focus on defining the needs and demographics of the target population, which in turn

should drive the expected outcomes of the intervention. As long as programs explicitly state the context of their evaluation and choose from among the core set of data elements, meaningful comparisons among existing programs should be feasible. While methods for collecting these metrics will depend upon resources and existing infrastructure, programs should aspire for rigor with objective sources when possible. When objective electronic data are not available, sites need creativity to determine the best way to retrieve the information, either from manual chart abstraction or navigator documentation. These recommendations are a first step toward adopting a minimal dataset for PED navigation programs, as has been done by other population based approaches to improving quality care.[54]

Navigation is emerging as an expected "standard" for cancer programs,[54] yet the literature has yet to provide consistent insight into activities or processes of navigation that are linked to favorable outcomes. We demonstrate here the growing body of knowledge regarding the impact of prevention and early detection navigation on cancer care would benefit from some thoughtful standardization. In keeping with recommendations from the 2001 IOM report to deliver patient-centered care that is timely, efficient and equitable[3] it is imperative that we evaluate the ability of PED patient navigation programs to realize that potential. Only then can we "apply evidence to health care delivery" as recommended. The responsibility for the analysis and synthesis of this medical evidence falls on all of us involved in the delivery of these services.

CONFLICT OF INTEREST DISCLOSURES

The authors made no disclosures.

REFERENCES

1. Edwards BK, Ward E, Kohler BA, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer*. 116:544-573. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19998273.
2. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med.* 2005; 353:1784-1792.
3. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
4. Elmore JG, Nakano CY, Linden HM, Reisch LM, Ayanian JZ, Larson EB. Racial inequities in the timing of breast cancer detection, diagnosis, and initiation of treatment. *Med Care.* 2005;43:141-148.
5. Mandelblatt JS, Yabroff KR, Kerner JF. Equitable access to cancer services: A review of barriers to quality care. *Cancer.* 1999;86:2378-2390.
6. [No Authors Listed]. A summary of the American Cancer Society report to the nation: cancer in the poor. *CA Cancer J Clin.* 1989;39:263-265.
7. Wells KJ, Battaglia TA, Dudley DJ, et al. Patient navigation: state of the art or is it science? *Cancer.*2008;113:1999-2010.
8. Freeman HP, Muth BJ, Kerner JF. Expanding access to cancer screening and clinical follow-up among the medically underserved. *Cancer Pract.* 1995;3:19-30.
9. Ell K, Padgett D, Vourlekis B, et al. Abnormal mammogram follow-up: a pilot study women with low income. *Cancer Pract.* 2002;10:130-138.
10. Ell K, Vourlekis B, Lee PJ, Xie B. Patient navigation and case management following an abnormal mammogram: a randomized clinical trial. *Prev Med.* 2007;44:26-33.
11. Ell K, Vourlekis B, Muderspach L, et al. Abnormal cervical screen follow-up among low-income Latinas: Project SAFe. *J Womens Health Gend Based Med.* 2002;11:639-651.
12. Battaglia TA, Roloff K, Posner MA, Freund KM. Improving follow-up to abnormal breast cancer screening in an urban population. A patient navigation intervention. *Cancer.*2007;109(suppl):359-367.
13. Bastani R, Mojica CM, Berman BA, Ganz PA. Low-income women with abnormal breast findings: results of a randomized trial to increase rates of diagnostic resolution. *Cancer Epidemiol Biomarkers Prev.* 2010;19:1927-1936.
14. Burhansstipanov L, Dignan MB, Schumacher A, Krebs LU, Alfonsi G, Apodaca CC. Breast screening navigator programs within three settings that assist underserved women. *J Cancer Educ.* 2010;25:247-252.
15. Burhansstipanov L, Dignan MB, Wound DB, Tenney M, Vigil G. Native American recruitment into breast cancer screening: the NAWWA project. *J Cancer Educ.* 2000;15:28-232.
16. American College of Surgeons Commission on Cancer. Cancer Program Standards (CPS) 2011 Project. Available at: http://www.facs.org/cancer/coc/cps2011.html accessed on December 6, 2010.
17. Oluwole SF, Ali AO, Adu A, et al. Impact of a cancer screening program on breast cancer stage at diagnosis in a medically underserved urban community. *J Am Coll Surg.* 2003;196:180-188.
18. Gabram SG, Lund MJ, Gardner J, et al. Effects of an outreach and internal navigation program on breast cancer diagnosis in an urban cancer center with a large African-American population. *Cancer.* 2008;113:602-607.
19. Dignan MB, Burhansstipanov L, Hariton J, et al. A comparison of two Native American Navigator formats: face-to-face and telephone. *Cancer Control.* 2005;12(suppl 2): 28-33.
20. Phillips CE, Rothstein JD, Beaver K, Sherman BJ, Freund KM, Battaglia TA. Patient navigation to increase mammography screening among inner city women. *J Gen Intern Med.* 2011;26:123-129.
21. Weber BE, Reilly BM. Enhancing mammography use in the inner city. A randomized trial of intensive case management. *Arch Intern Med.* 1997;157:2345-2349.
22. Jandorf L, Gutierrez Y, Lopez J, Christie J, Itzkowitz SH. Use of a patient navigator to increase colorectal cancer screening in an urban neighborhood health clinic. *J Urban Health.* 2005;82:216-224.

23. Lasser KE, Murillo J, Medlin E, et al. A multilevel intervention to promote colorectal cancer screening among community health center patients: results of a pilot study. *BMC Fam Pract.*2009;10:37.
24. Nash D, Azeez S, Vlahov D, Schori M. Evaluation of an intervention to increase screening colonoscopy in an urban public hospital setting. *J Urban Health.* 2006;83:231-243.
25. Chen LA, Santos S, Jandorf L, et al. A program to enhance completion of screening colonoscopy among urban minorities. *Clin Gastroenterol Hepatol.* 2008;6:443-450.
26. Paskett E, Tatum C, Rushing J, et al. Randomized trial of an intervention to improve mammography utilization among a triracial rural population of women. *J Natl Cancer Inst.* 2006;98:1226-1237.
27. Han HR, Lee H, Kim MT, Kim KB. Tailored lay health worker intervention improves breast cancer screening outcomes in non-adherent Korean-American women. *Health Educ Res.* 2009;24:318-329.
28. Wang X, Fang C, Tan Y, Liu A, Ma GX. Evidence-based intervention to reduce access barriers to cervical cancer screening among underserved Chinese American women. *J Womens Health (Larchmt).* 2010;19:463-469.
29. Fernandez ME, Gonzales A, Tortolero-Luna G, et al. Effectiveness of Cultivando la Salud: a breast and cervical cancer screening promotion program for low-income Hispanic women. *Am J Public Health.* 2009;99:936-943.
30. Ford ME, Havstad S, Vernon SW, et al. Enhancing adherence among older African American men enrolled in a longitudinal cancer screening trial. *Gerontologist.* 2006;46:545-550.
31. Clark CR, Baril N, Kunicki M, et al. Addressing social determinants of health to improve access to early breast cancer detection: results of the Boston REACH 2010 Breast and Cervical Cancer Coalition Women's Health Demonstration Project. *J Womens Health (Larchmt).* 2009;18:677-690.
32. Percac-Lima S, Grant RW, Green AR, et al. A culturally tailored navigator program for colorectal cancer screening in a community health center: a randomized, controlled trial. *J Gen Intern Med.* 2009;24:211-217.
33. Dietrich AJ, Tobin JN, Cassells A, et al. Telephone care management to improve cancer screening among low-income women: a randomized, controlled trial. *Ann Intern Med.* 2006;144:563-571.
34. Christie J, Itzkowitz S, Lihau-Nkanza I, Castillo A, Redd W, Jandorf L. A randomized controlled trial using patient navigation to increase colonoscopy screening among low-income minorities. *J Natl Med Assoc.* 2008;100:278-284.
35. Ma GX, Shive S, Tan Y, et al. Community-based colorectal cancer intervention in underserved Korean Americans. *Cancer Epidemiol.* 2009;33:381-386.
36. Myers RE, Hyslop T, Sifri R, et al. Tailored navigation in colorectal cancer screening. *Med Care* 2008;46(suppl 1): S123-S131.
37. Ferrante JM, Chen PH, Kim S. The effect of patient navigation on time to diagnosis, anxiety, and satisfaction in urban minority women with abnormal mammograms: a randomized controlled trial. *J Urban Health.* 2008;85:114-124.
38. Palmieri FM, DePeri ER, Mincey BA, et al. Comprehensive diagnostic program for medically underserved women with abnormal breast screening evaluations in an urban population. *Mayo Clin Proc.* 2009;84:317-322.
39. Psooy BJ, Schreuer D, Borgaonkar J, Caines JS. Patient navigation: improving timeliness in the diagnosis of breast abnormalities. *Can Assoc Radiol J.* 2004;55:145-150.
40. Fouad M, Wynn T, Martin M, Partridge E. Patient navigation pilot project: results from the Community Health Advisors in Action Program (CHAAP). *Ethn Dis.* 2010;20:155-161.
41. Freund KM, Battaglia TA, Calhoun E, et al. National Cancer Institute Patient Navigation Research Program: methods, protocol, and measures. *Cancer.* 2008;113:3391-3399.
42. Lin CJ, Schwaderer KA, Morgenlander KH, et al. Factors associated with patient navigators' time spent on reducing barriers to cancer treatment. *J Natl Med Assoc.* 2008;100:1290-1297.
43. Rauscher GH, Johnson TP, Cho YI, Walk JA. Accuracy of self-reported cancer-screening histories: a meta-analysis. *Cancer Epidemiol Biomarkers Prev.* 2008;17:748-757.
44. Ponce NA, Lavarreda SA, Yen W, Brown ER, DiSogra C, Satter DE. The California Health Interview Survey 2001: translation of a major survey for California's multiethnic population. *Public Health Rep.* 2004;119:388-395.
45. Parker VA, Clark JA, Leyson J, et al. Patient navigation: development of a protocol for describing what navigators do. *Health Serv Res.* 2010;45:514-531.
46. Campbell SM, Roland MO, Shekelle PG, Cantrill JA, Buetow SA, Cragg DK. Development of review criteria for assessing the quality of management of stable angina, adult asthma, and non-insulin dependent diabetes mellitus in general practice. *Qual Health Care.* 1999;8:6-15.
47. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q.* 1966;44(suppl):166-206.
48. Hearnshaw HM, Harker RM, Cheater FM, Baker RH, Grimshaw GM. Expert consensus on the desirable characteristics of review criteria for improvement of health care quality. *Qual Health Care.* 2001;10:173-178.
49. Brook RH, McGlynn EA, Shekelle PG. Defining and measuring quality of care: a perspective from US researchers. *Int J Qual Health Care.* 2000;12:281-295.
50. Crombie IK, Davies HT. Beyond health outcomes: the advantages of measuring process. *J Eval Clin Pract.* 1998;4:31-38.
51. Rubin HR, Pronovost P, Diette GB. The advantages and disadvantages of process-based measures of health care quality. *Int J Qual Health Care.* 2001;13:469-474.
52. Donabedian A. The quality of care: How can it be assessed? *JAMA.* 1988;260:1743-1748.
53. Campbell HS, Ossip-Klein D, Bailey L, Saul J. Minimal dataset for quitlines: a best practice. *Tob Control.*2007;16 (suppl 1):i16-i20.
54. American College of Surgeons Commission on Cancer. Cancer Program Accreditation. Available at: http://www.facs.org/cancer/coc/sarguide.html accessed on January 31, 2011.
55. American Joint Committee on Cancer. The A-JCC e-Staging Tool. Available at: https://www.ajcce-staging.com/staging/view/cancerStage accessed on January 3, 2011.
56. Sobin L, Gospodarowicz M, Wittekind C, eds. TNM Classification of Malignant Tumors. 7th ed. Oxford, England: Wiley-Blackwell; 2009.

Improving Disposition Outcomes for Patients in a Geriatric Skilled Nursing Facility

Randi E. Berkowitz, MD,† Richard N. Jones, ScD,*† Ron Rieder,* Margaret Bryan,* Robert Schreiber, MD,*† Sharon Verney, RN/NP,* and Michael K. Paasche-Orlow, MD, MA, MPH‡*

OBJECTIVES: To evaluate an intervention to improve discharge disposition from a skilled nursing unit (SNU).

DESIGN: Historical control comparison of discharge disposition before and after implementation.

SETTING: Fifty-bed SNU.

PARTICIPANTS: All patients admitted from acute care hospitals to a SNU between June 2008 and May 2010.

INTERVENTION: Physician admission procedures were standardized using a template, patients with three or more hospital admissions over the prior 6 months received palliative care consultations, and multidisciplinary root-cause analysis conferences for patients transferred back to the hospital acutely were conducted bimonthly to identify problems and improve processes of care.

MEASUREMENTS: Patients' discharge disposition (i.e., acute care, long-term care, home, or death) before and after implementation were compared.

RESULTS: Discharge dispositions were determined for all 1,725 patients admitted during the study; 862 patients before (June–May 2008) and 863 during (June 2009–May 2010) the intervention. Discharge dispositions were significantly differently distributed across the two periods ($P = .03$). Readmission to acute care declined (from 16.5% to 13.3%, a nearly 20% decline). Multivariable logistic regression, controlling for age, sex, and case-mix index and adjusting for clustering due to repeated admissions of individual patients, suggests that, during the intervention period, patients were more likely than during the baseline period to die on the unit in accordance with their wishes than to be transferred out to the hospital (odds ratio = 2.45, 95% confidence interval = 1.09–5.5).

CONCLUSION: Interventions such as the ones implemented can lead to fewer hospital transfers for SNUs. **J Am Geriatr Soc 59:1130–1136, 2011.**

Key words: rehospitalization; skilled nursing unit; palliative care; multidisciplinary team

One in five Medicare beneficiaries was rehospitalized within 30 days of hospital discharge in 2004, at an estimated cost of $17.4 billion.[1] Hospitalized patients admitted to a skilled nursing facility (SNF) have a high rate of early unplanned rehospitalization.[2] In 2006, the national rate for patients discharged to a SNF who were rehospitalized directly from the SNF or within 2 days of discharge from the SNF was 23.5%. Two reasons to believe that a fair amount of these events are likely to be avoidable are the high prevalence of preventable diagnoses and significant geographic variation.[3] For example, whereas patients in Utah discharged to SNF had a rehospitalization rate of 15.1%, patients in Louisiana had a rate of 28.2%.[2]

Hospitals are currently required to report readmission rates, but few SNFs use repeat hospitalizations as a measure of quality of care. Because SNFs typically serve patients who are admitted to and discharged from multiple hospitals—and SNF administrators may not have access to these data—SNF administrators cannot generally determine the rate of readmission for their patients once they have been discharged to the community. SNFs do have access to Minimum Data Set (MDS) data to follow the number of patients they are sending out acutely to the hospital, but this is currently not a required quality indicator.

Many factors contribute to rehospitalization risk. Risk factors include prior recent hospitalization, specific diagnoses (e.g., congestive heart failure), and indices such as carbon dioxide levels for patients with chronic obstructive pulmonary disease, renal function, and other clinical parameters.[4–8] Clinical instability, lack of medication reconciliation, depression, and multiple other factors also

From the *Hebrew SeniorLife, Boston, Massachusetts; †Division of Gerontology Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, Massachusetts; and ‡Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts.

Meeting Presentation: American Medical Directors Association, March 2011.

Address correspondence to Randi Berkowitz, 1200 Centre Street, Roslindale, MA 02131. E-mail: rberkowitz@hsl.harvard.edu

DOI: 10.1111/j.1532-5415.2011.03417.x

0002-8614/11/$15.00

contribute to rehospitalization risk.[9–17] Geriatric assessment, nurse practitioner involvement, and type of facility have been found to be associated with a lower rate of readmissions from SNFs.[18–21]

This project was a prospective quasi-experimental trial to change discharge dispositions for patients on a skilled nursing unit (recuperative services unit, RSU). The intervention included three elements: Physician admission procedures were standardized with a template, which included care guidelines for common geriatric syndromes, a template for medicine reconciliation, a standardized goals of care discussion, and a question of how many times the patient had been hospitalized over the past 6 months; patients with more than three hospital admissions over the prior 6 months received automatic palliative care consultation; and multidisciplinary conferences were conducted every 2 weeks examining the care of patients acutely transferred to the hospital to identify problems and improve processes of care. The distribution of discharge dispositions before the intervention and after initiation of the intervention were compared to evaluate the efficacy of the program.

METHODS

The Hebrew SeniorLife (HSL) institutional review board approved the project, and an advisory committee composed of nurses, secretarial staff, aides, therapists, nurse practitioners, doctors, administrators from HSL and an acute care hospital, and a daughter of a patient who had experienced repeat hospitalizations was convened and met before implementation of the intervention and biannually to guide the project. The main outcome of interest was the distribution of discharge dispositions, including readmission, transfer to long-term care facility, discharge to home, and death. Readmission includes people transferred directly from the RSU to an acute care hospital or psychiatric unit, transferred to an acute care hospital or psychiatric unit at the time of a physician office visit during the time they are an RSU patient, and transferred to acute care or psychiatric unit from a dialysis unit during the time they are an RSU patient. The HSL medical care review committee reviewed all deaths and acute care transfers at HSL to ascertain potential errors and avoidable causes and to determine whether deaths were expected.

The baseline distribution of discharge dispositions, before initiation of the intervention, included data from all patients admitted to the RSU from June 2008 to May 2009. The intervention began in June 2009, and data for the first 12 months of the intervention were used for the current analyses. Specifically, the intervention period included final disposition data from all patients admitted to the RSU from June 2009 to May 2010. Administrative data were used to determine the discharge disposition for each person. In addition, initial information was collected from the MDS assessment, including age, sex, race, ethnicity, functional level (activity of daily living long-form scale), cognitive performance level (Cognitive Performance Scale), pain index, depression rating scale, and case-mix index (CMI). All patients were on Medicare Part A or managed care. The RSU is part of HSL, a nonprofit institution with a closed medical staff, and is located in a facility that also includes a 400-bed long-term care hospital. Doctors and nurse practitioners are on site until 6 p.m. Monday to Friday and for rounds on weekends and holidays, and physicians are on-call by telephone at other times.

In 2008, a standardized template was developed with input from all members of the HSL Department of Medicine and important nursing, administrative, and social service leaders. The American Medical Directors admission history and physical template was used, and care guidelines for common geriatric syndromes, medicine reconciliation, goals of care, and a question of how many times the patient had been hospitalized over the past 6 months were included (Appendix 1). The advance directives section included a discussion about whether the patient or healthcare proxy would want subsequent hospitalizations if the patient's condition deteriorated while on the SNF. To determine fidelity to this aspect of the intervention protocol, a random sample of 40 patients' charts was surveyed to determine whether the admitting attending used the template.

If a patient had had three or more hospitalizations (including the hospitalization immediately preceding the current SNF admission) in the past 6 months, a palliative care consultation was obtained with patient consent to identify realistic goals of care and address barriers to discharge home. The palliative care team was composed of a physician board certified in palliative care, a geriatric nurse, a geriatric social worker, and a chaplain. All members of the team were also encouraged to ask for a palliative care consultation if they believed there was discordance between the team, the family, and the patient's expectations for progress. The objective of the palliative care consultation was to determine whether rehospitalization was consistent with the patient's goals of care or if worsening symptoms would best be managed in the SNF, long-term care, or at home.

Team Improvement for the Patient and Safety (TIPS) conferences were held twice a month for 30 minutes, starting in June 2009, to examine the root causes of rehospitalization events. Nurses, nursing aides, physicians, therapists, social workers, and a nursing home administrator attended sessions. Meeting times were varied to ensure that night and evening staff were included, and aides were compensated for attending TIPS conferences after their shifts had ended.

At TIPS conferences, selected cases of rehospitalization that were deemed to have been potentially avoidable were reviewed to identify ways in which the team could have operated more effectively. Before the TIPS conference, physicians called the readmitting hospital and spoke with the hospital care team to gain insights into problems that might have been missed on the SNF. According to the specific causes identified, additional information would be sought, and additional staff or outside experts were invited to participate in the TIPS session. During the course of the year, representatives from security, maintenance, home care agencies, inpatient and outpatient pharmacies, information technology, psychiatry, recreation therapy, dietary, admissions, covering physicians, palliative care, respiratory therapy, families, and laboratory staff were included in the TIPS conference. Attendance was measured. An email list of all direct care staff was created, and a "lessons learned" email was shared after each meeting.

Components of the intervention are summarized in Table 1.

Table 1. Interventions

Physician history and physical template with goals-of-care discussion including code status, number recent admissions in past 6 months, and whether repeat hospitalization is consistent with patient's wishes
Palliative care consult with patient consent if more than three hospitalizations over past 6 months
Physician call to hospital on unplanned discharges to determine whether diagnosis missed in skilled nursing facility
Multidisciplinary conferences every 2 weeks to review cases of unplanned discharges to identify and fix system failures

Statistical Methods

The distribution of demographic and clinical characteristics of patients and the pattern of discharges were compared between the two study periods, and the hypothesis that the two samples represented random samples from the same population was tested using simple bivariate tests (analysis of variance, chi-square) (Table 2).[22–25] Because some patients were represented multiple times in the data, with repeat RSU admissions, and straddled study periods, typical linear regression model assumptions of independence of observations are not consistent with these data. This was addressed, and the differences in the distribution of discharges across the two study periods were formally tested using multinomial logistic regression modeling with robust standard errors controlling for clustering on individual residents. For both study periods, 6% of patients were missing covariate data, because they did not have an MDS assessment before discharge. Missing data were handled with multiple imputation methods, using the discharge outcome as the predictor in the multiple imputation models. Regression models used standard methods for pooling results over multiple estimations.[26] Parameter estimates were obtained using Stata software (version 10.1, Stata Corp., College Station, TX).

Table 2. Patient Characteristics for the Hebrew SeniorLife Recuperative Services Unit

Characteristic	Before Start of Intervention (June 2008 to May 2009), n = 862	After Start of the Intervention (June 2009 to May 2010), n = 863	Significance Test
Age, mean ± SD	82.7 ± 9.1	82.0 ± 9.8	$F = 2.52$; $P = .11$
Sex, n (%)			$\chi^2 = 2.30$; $P = .13$
Male	268 (33.1)	240 (29.6)	
Female	542 (66.9)	571 (70.4)	
Length of stay, days, mean ± SD	14.9 ± 12.2	14.6 ± 12.9	$F = 0.35$; $P = .55$
Race or ethnicity, n (%)			$\chi^2 = 5.70$; $P = .13$
Asian or other Pacific Islander	0 (0.0)	3 (0.4)	
Black, not Hispanic	50 (6.2)	67 (8.3)	
White, not Hispanic	743 (92.0)	725 (89.6)	
Hispanic	15 (1.9)	14 (1.7)	
Activity of daily living Long-Form Scale score (range 0–27), mean ± SD	14.7 ± 4.3	14.6 ± 4.1	$F = 0.04$; $P = .85$
Cognitive Performance Scale score, n (%)			$\chi^2 = 5.40$; $P = .37$
Intact	486 (60.0)	461 (56.8)	
Borderline intact	140 (17.3)	143 (17.6)	
Mild impairment	84 (10.4)	97 (12.0)	
Moderate impairment	84 (10.4)	98 (12.1)	
Moderate to severe impairment	11 (1.4)	11 (1.4)	
Severe impairment	5 (0.6)	1 (0.1)	
Depression rating scale score, mean ± SD	0.3 ± 0.7	0.2 ± 0.6	$F = 5.34$; $P = .02$
Pain index, n (%)			$\chi^2 = 15.40$; $P = .002$
No pain	220 (27.2)	171 (21.1)	
Pain less than daily	221 (27.3)	280 (34.5)	
Daily mild to moderate	274 (33.8)	284 (35.0)	
Daily excruciating	95 (11.7)	76 (9.4)	
Case Mix Index, mean ± SD	1.4 ± 0.3	1.4 ± 0.3	$F = 2.61$; $P = .11$
Discharge disposition, n (%)			$\chi^2 = 8.70$; $P = .03$
Community	591 (68.6)	630 (73.0)	χ^2 cont = 1.2
Died	10 (1.2)	19 (2.2)	χ^2 cont = 2.8
Another facility	119 (13.8)	99 (11.5)	χ^2 cont = 1.9
Hospitalization	142 (16.5)	115 (13.3)	χ^2 cont = 2.9

SD = standard deviation; χ^2 cont = contribution to overall chi-square for the row-wise comparison.

RESULTS

Eight hundred sixty-two people were admitted to the RSU in the 12 months before the intervention, and 863 were admitted during the 12 months of the intervention. Patient age, sex, race, functional status, cognitive level at baseline, case-mix adjustment, and length of stay did not differ significantly between the control and intervention years (Table 2). In 2007, the latest year available, the CMI for the unit was 1.21, whereas the national average for hospital based facilities was 0.94 ± 0.19.

In each year, 52 patients had incomplete MDS assessment (were discharged or died before MDS assessment). Seventy-nine percent of the patients had one admission to the RSU during the 2-year interval, 16% had two, and 5% had three or more (maximum of 6). Physicians used the standardized admission assessment template in 35 of 40 (87.5%) audited charts. All patients had physician orders documenting code status. There were 55 palliative care consultations in the control year and 116 in the intervention year.

During the course of the intervention period, 22 TIPS conferences were held; of staff on duty at the time of the conference, there was an average attendance rate at TIPS conferences of 81%.

Discharge dispositions differed significantly between years ($P = .03$), with the rate of rehospitalization declining from 16.5% to 13.3%, a drop of 19.4% (Table 3). Discharges to home increased from 68.6% to 73.0%, deaths on the RSU increased from 1.2% to 2.2%, and discharges to long-term care fell from 13.8% to 11.5%. The medical care review committee judged all deaths to be expected and consistent with patient wishes.

Multivariable logistic regression, controlling for age, sex, case-mix index, and repeated admissions of individual patients, indicated that patients were more likely to die on the unit than be transferred out to the hospital during the intervention than during the baseline period (odds ratio = 2.45, 95% confidence interval = 1.09–5.5).

DISCUSSION

After implementing the three-pronged intervention, there was a change in discharge disposition from the SNF, with a decline in discharges to acute and long-term care and increases in discharges to home and palliative care deaths on the unit. Two components of the intervention—standardized admission assessments and multidisciplinary conferences discussing root-cause analysis for patients acutely transferred back to the hospital—were conducted with existing staff. Many SNFs could embed similar programs within their current care processes. Instituting this program may require additional resources such as time to institute the admission template, palliative care services, and staff time for TIPS conferences. Teams from hospice organizations that are already embedded in many long-term care facilities could aid organizations without a palliative care service.

The authors feel that the change in discharge disposition observed between the two periods reflects a true improvement in patient outcomes, although some caution is required when interpreting these results. Specifically, a lower acute transfer rate probably reflects better processes of care in the SNF, but there is no criterion standard to evaluate physician judgments regarding the appropriateness to transfer or not transfer patients to the hospital. In addition, all deaths on the unit were concordant with patient wishes, another important indicator that the observations reflect an improvement in patient care.

This model can be disseminated. Organizations considering projects to improve care transitions can compare their population with the current study population using a resource developed by the Shaping Long Term Care in America Project on their Web site http://www.ltcfocus.org.[27]

Table 3. Discharge Status: Multivariable Multinomial Logistic Regression Modeling

	Odds Ratio			
Outcome	**Community**	**Hospitalization**	**Facility**	**Died**
Adjusted for clustering on individual only				
Died	1.78	2.35*	2.28*	—
Facility	0.78	1.03	—	0.44*
Hospitalization	0.76	—	0.97	0.43*
Community	—	1.32	1.28	0.56
Adjusted for clustering on individual and covariates				
Died	1.91	2.45*	2.42*	—
Facility	0.79	1.01	—	0.41*
Hospitalization	0.78	—	0.99	0.41*
Community	—	1.28	1.27	0.52
Test time effect = 0 (χ^2 (degrees of freedom))				8.76 (3)*
F-test $P > \chi^2$				0.033

* $P < .05$, test that individual level regression parameter is significantly different from the null. Covariates include age, sex, functional level (activity of daily living long-form scale), cognitive performance level (Cognitive Performance Scale), and case-mix index. *Note on interpreting parameter estimates*: 1.91 is the increase in the odds of dying versus being discharged to the community, comparing persons visiting the recuperative services unit in the intervention period versus the baseline period holding other variables in the model constant. Significance tests in all models are estimated using robust standard errors adjusting for clustering on individual.

Because the population in the current study was in the top 10% of acuity based on national CMI data, other SNFs with lower acuity may expect different results.

This study has several limitations that should be discussed. First, it was not possible to separate the effect of the three components of the intervention, partly because of limited details collected regarding the effect of each component of the intervention and the nature of the study design. For example, it is unknown whether the template improved the rate of guideline-concordant care for geriatric syndromes. Issues of transitions of care are multifactorial and need systematic response from the beginning to the end of the care process. The intervention was designed to promote the importance of patient's goals of care and to help staff see transitions of care as an important part of their work product. Attitudes of culture change are currently being studied, and the Agency for Healthcare Research and Quality Long-Term Care Patient Survey is being used to quantify these changes.[28] Further studies would be needed to delineate the relative contribution of each aspect of the intervention.

The second limitation to address is generalizability to other SNFs. Further studies, for example, would be needed to see how to adapt the intervention for facilities without an onsite medical staff. In addition, the baseline hospital readmission rate of 16.5% on the RSU is already particularly low. The average 30-day readmission rates for people who have been in a SNF in Massachusetts is approximately 22% for patients going from home to hospital to SNF and 28% for SNF patients who are hospitalized and discharged to a SNF.[2] It is likely that the low transfer rates at the RSU may reflect the ability of onsite medical staff to assess acute medical conditions quickly. Similar projects might have an even larger effect in facilities with higher baseline rates of acute transfers.

A third limitation of this study is that it was not a randomized trial. As an intervention that aimed to influence the interaction between staff and patients and to improve organizational attention to care transitions, randomization could not be done at the patient level because of the likelihood of contamination. Cluster randomization according to site of care was outside the scope of the current project but would be feasible with adequate funding. Despite the limitations of the quasi-experimental design, the fact that the case-mix index and other patient characteristics were unchanged from the baseline period to the intervention period are reasons to feel confident that the observed improvement in discharge dispositions reflects a true intervention effect.

A fourth limitation of this study is that data for what happens to people after they are transferred from the RSU were not available. Although complete data on discharge disposition were available, data on subsequent care transitions were not. People who are discharged to their homes may then be admitted to various hospitals or facilities, and there is no easy way to track these events. An important development would be for states to facilitate data collection and analysis of readmission rates to enable facilities to monitor the effectiveness of their discharge planning. Until facilities have access to such data, SNFs should be required to report risk-adjusted acute transfer rates.

A final limitation is that the fidelity of the intervention was not fully monitored. Of the 863 patient admissions that occurred in the intervention period, 40 were monitored, and evidence of adherence to the intervention was found in 87.5%. It is unclear how much additional benefit a higher rate of adoption of the intervention activities might have yielded.

The three components of the intervention—the standardized admissions template, palliative care consultations, and the TIPS conference—represent different types of activities that were designed to improve transitions of care. Order sets have been shown to promote quality of care in various settings but have not been evaluated in SNFs.[29–31] The template includes triggers to aid goals-of-care discussions and evaluation of the rehospitalization rate to trigger consultation by the palliative care team. The purpose of discussing goals of care and of having the palliative care team involved is to ensure that the care delivered is consistent with patients' wishes. Rehabilitation staff are frequently focused on restoring a patient's function and are not necessarily equipped to help families and patients recognize when there may be a permanent decline in function. The palliative care team not only educated families and patients, but also coached nursing and therapy staff for symptom management. Although these activities involved important members of the care team, the TIPS conference series was designed to include a broad representation of staff, allowing for ongoing organizational emphasis on the importance of transitions of care in a manner that highlighted opportunities for improvement.

During the intervention period, sick patients were kept on the unit if they did not wish to be rehospitalized. This potentially increased the cost of providing care. Because SNFs are not reimbursed for the extra care such patients require, it is easy to see why patients are routinely sent back to acute care settings simply for lack of staffing at the SNF. Financial incentives should promote avoidance of unneeded rehospitalizations. It is hoped that the bundled payment scheme of the Accountable Care Act will provide physicians and hospitals with adequate incentives to coordinate care for patients at SNFs.

ACKNOWLEDGMENTS

We thank Julie Harris, Ann Masterson, Scott Ariel, the palliative care team, Jonathan Regosin, Fran Hinckley, Susan McCurley, Terry O'Malley, Eli Shea, Lou Woolf, Len Fishman, Richard Della Penna, Kyle Allen, and Patricia Leddy for their advice and encouragement. The Shaping Long-Term Care in American Project provided data for national CMI. We also thank every aide, nurse, therapist, provider, and social worker who attended to the needs of our patients at HSL.

Conflict of Interest: None of the authors report any personal, potential, or financial conflicts of interest.

This project was funded by the Practice Change Fellowship, Denver, Colorado and Hebrew SeniorLife.

Author Contributions: Randi Berkowitz: primary investigator and designer of project and TIPS conferences and preparation of manuscript. Richard N. Jones: statistician and preparation of manuscript. Ron Reider: information technology consultant. Margaret Bryan: database manager.

Robert Schreiber: helped with formulating project and with TIPS conferences. Sharon Verney: case reviews and TIPS conferences. Michael K Paasche-Orlow: analysis and interpretation of data and preparation of manuscript.

Sponsor's Role: The Practice Change Fellowship supported 20% of Dr. Berkowitz's salary to complete a performance improvement project and had no role in the data or the manuscript preparation. The fellowship also supported leadership training and mentoring.

HSL donated 20% of Dr. Berkowitz's salary with in-kind support to lead and plan TIPS conferences and data management in addition to supporting the salary of the palliative care team members.

REFERENCES

1. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med 2009;360:1418–1428.
2. Mor V, Intrator O, Feng Z et al. The revolving door of rehospitalization from skilled nursing facilities. Health Aff (Millwood) 2010;29:57–64.
3. Donelan-McCall NET, Eilerson T, Fish R et al. Small Patient Population and Low Frequency Event Effects on the Stability of SNF Quality Measures. Washington, DC: Medicare Payment Advisory Commission, 2006.
4. Bohannon RW, Maljanian R, Ferullo J et al. Mortality and readmission of the elderly one year after hospitalization for pneumonia. Aging Clin Exp Res 2004;16:22–25.
5. Almagro P, Barreiro B, Ochoa de Echaguen A et al. Risk factors for hospital readmission in patients with chronic obstructive pulmonary disease. Respiration 2006;73:311–317.
6. Silverstein MD, Qin H, Mercer SQ et al. Risk factors for 30-day hospital readmission in patients>/=65 years of age. Proc (Bayl Univ Med Cent) 2008;21:363–372.
7. Tamhane U, Voytas J, Aboufakher R et al. Do hemoglobin and creatinine clearance affect hospital readmission rates from a skilled nursing facility heart failure rehabilitation unit? J Am Med Dir Assoc 2008;9:194–198.
8. Novotny NL, Anderson MA. Prediction of early readmission in medical inpatients using the Probability of Repeated Admission instrument. Nurs Res 2008;57:406–415.
9. Vecchiarino P, Bohannon RW, Ferullo J et al. Short-term outcomes and their predictors for patients hospitalized with community-acquired pneumonia. Heart Lung 2004;33:301–307.
10. Moore C, Wisnivesky J, Williams S et al. Medical errors related to discontinuity of care from an inpatient to an outpatient setting. J Gen Intern Med 2003;18:646–651.
11. Campbell SE, Seymour DG, Primrose WR. A systematic literature review of factors affecting outcome in older medical patients admitted to hospital. Age Ageing 2004;33:110–115.
12. Muzzarelli S, Leibundgut G, Maeder MT et al. Predictors of early readmission or death in elderly patients with heart failure. Am Heart J 2010;160:308–314.
13. Babayan ZV, McNamara RL, Nagajothi N et al. Predictors of cause-specific hospital readmission in patients with heart failure. Clin Cardiol 2003;26:411–418.
14. Tuppin P, Neumann A, Danchin N et al. Evidence-based pharmacotherapy after myocardial infarction in France: Adherence-associated factors and relationship with 30-month mortality and rehospitalization. Arch Cardiovasc Dis 2010;103:363–375.
15. Mitchell SE, Paasche-Orlow MK, Forsythe SR et al. Post-discharge hospital utilization among adult medical inpatients with depressive symptoms. J Hosp Med 2010;5:378–384.
16. Capelastegui A, Espana Yandiola PP et al. Predictors of short-term rehospitalization following discharge of patients hospitalized with community-acquired pneumonia. Chest 2009;136:1079–1085.
17. Jasti H, Mortensen EM, Obrosky DS et al. Causes and risk factors for rehospitalization of patients hospitalized with community-acquired pneumonia. Clin Infect Dis 2008;46:550–556.
18. Burl JB, Bonner A, Rao M et al. Geriatric nurse practitioners in long-term care: Demonstration of effectiveness in managed care. J Am Geriatr Soc 1998; 46:506–510.
19. Caplan GA, Williams AJ, Daly B et al. A randomized, controlled trial of comprehensive geriatric assessment and multidisciplinary intervention after discharge of elderly from the emergency department—the DEED II study. J Am Geriatr Soc 2004;52:1417–1423.
20. Harris R, Wilson-Barnett J, Griffiths P. Effectiveness of nursing-led inpatient care for patients with post-acute health care needs: Secondary data analysis from a programme of randomized controlled trials. J Eval Clin Pract 2007;13:198–205.
21. Kauh B, Polak T, Hazelett S et al. A pilot study: Post-acute geriatric rehabilitation versus usual care in skilled nursing facilities. J Am Med Dir Assoc 2005;6:321–326.
22. Morris JN, Fries BE, Morris SA. Scaling ADLs within the MDS. J Gerontol A Biol Sci Med Sci 1999;54A:M546–M553.
23. Morris JN, Fries BE, Mehr DR et al. MDS cognitive performance scale. J Gerontol 1994;49:M174–M182.
24. Proctor WR, Hirdes JP. Pain and cognitive status among nursing home residents in Canada. Pain Res Manage 2001;6:119–125.
25. Fries BE, Schneider DP, Foley WJ et al. Refining a case-mix measure for nursing homes: Resource Utilization Groups (RUG-III). Med Care 1994;32:668–685.
26. Rodriguez-Artalejo F, Guallar-Castillon P, Pascual CR et al. Health-related quality of life as a predictor of hospital readmission and death among patients with heart failure. Arch Intern Med 2005;165:1274–1279.
27. Long-term Care: Facts on Care in the US [on-line]. Available at http://www.ltcfocus.org Accessed January 17, 2011.
28. Winograd CH, Gerety MB, Chung M et al. Screening for frailty: Criteria and predictors of outcomes. J Am Geriatr Soc 1991;39:778–784.
29. O'Connor CM, Abraham WT, Albert NM et al. Predictors of mortality after discharge in patients hospitalized with heart failure: An analysis from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF). Am Heart J 2008;156:662–673.
30. Ballard DJ, Ogola G, Fleming NS et al. Impact of a standardized heart failure order set on mortality, readmission, and quality and costs of care. Int J Qual Health Care 2010;22:437–444.
31. Bekmezian A, Chung PJ, Yazdani S. Standardized admission order set improves perceived quality of pediatric inpatient care. J Hosp Med 2009;4: 90–96.

APPENDIX 1

Admitting history and physical

Recuperative service unit
Date:
Referring hospital:
PCP: Telephone/fax:
SPEC./Surgeons: Telephone/fax:
Healthcare proxy phone number and name:
The patient is admitted to the RSU for:
Short-term rehabilitation:
History of present illness
Past medical history
Medications
Home medications:
Hospital medications changed from home:
Med reconciliation at Hebrew SeniorLife:
Family history:
No known drug allergies
Social history
Lives with:
Code status:
Services at home:
Alcohol use:
Tobacco use:
Spiritual/religious:
Infection control:
Contact Precautions [*]
Strict Precautions [*]
Droplet Precautions [*]
Neutropenic Precautions [*]
Immunization dates:
Influenza vaccine:

Pneumovax:

Foreign bodies [such ostomy, foley, ivs, CPAP, pacemaker, pessary]

Functional history

Ambulates independently

Transfers independently

Eats independently

Dresses independently

Toilets independently

Review of systems

General: Chronic pain negative, recent weight loss negative, overall decline negative, fatigue negative

Skin: Itching negative, new skin lesions negative, rash negative

Eyes: Visual changes negative, glasses negative, legal blindness negative, irritation redness negative

ENT: Hearing loss negative, difficulty chewing negative, difficulty swallowing negative, difficulty speaking negative, hoarseness negative, sore throat negative, ear pain negative

Respiratory: Shortness of breath negative, dyspnea on exertion, negative cough, negative hemoptysis

Cardiovascular system: Chest pain negative, palpitations negative, orthopnea negative, edema negative, claudication negative

Endocrine: Polydipsia negative, polyuria negative

Hematologic: Easy bruising negative

Gastrointestinal: Heartburn negative, abdominal pain negative, constipation negative, diarrhea negative, blood in stools negative, incontinence negative

Genitourinary: Nocturia negative, frequency negative, urgency negative, burning pain negative, hematuria negative, incontinence negative

Musculoskeletal: Joint pain negative, straight swelling negative, muscle pain negative, back pain negative

Neurological: Confusion negative, headache negative, dizziness negative, falls negative, gait disorder negative, numbness negative, weakness negative, tremor negative

Psychiatric: Memory loss negative, anxiety negative, depression negative, sleep disorder negative, delusions negative, hallucinations negative, agitation negative

Physical exam

Well-nourished, no apparent distress

Skin: with good turgor, no pressure ulcers, no rashes

Head: normocephalic, atraumatic

Eyes: PERRLA no nystagmus normal sclerae

HEENT: normal hearing, no sinus tenderness, oropharynx negative, good dentition, no lymphadenopathy

Neck: normal range of motion, no carotid bruits, thyroid negative

Chest: kyphotic, clear to auscultation. No rubs, rales, rhonchi, wheezes

Heart: no murmurs, normal S1-S2, no rubs or gallops

Peripheral vascular: 2+pulses

Breasts: no nipple discharge, no masses, no axillary adenopathy

Abdomen: nondistended, nontender, soft, positive bowel signs, no organomegaly, no rebound, no guarding

GU: negative

Extremities: no clubbing, cyanosis or edema. No contractures, no joint effusions, osteoarthritis changes

Neurologic: alert and oriented x3, cranials intact sensation, motor grossly normal

Gait: able to rise from a chair

Mini-Cog:

3 Item Recall Score: [ ]/3

Clock Draw Score: [ ]/1

Mini-Cog Score Total Score: [*]

Evidence of Confusion: yes no (if yes proceed with CAM assessment)

CAM Score: [*]

Laboratory data

Date:

Source: [ ] hospital [ ] admit

H/H: MCV: WBC: platelets:

BUN/Creatinine:

Na: K: CI: CO2:

Other:

ASSESSMENT/PLAN

BARRIERS TO DISCHARGE

Estimated RSU length of stay: [*] weeks time.

Rehabilitation potential: [*].

RSU goals: Increase strength and safety, stabilize and improve medical condition, prevent pain, prevent pressure sores and delirium, and increase independence in ADLs.

Diet: [*].

Physical therapy: Will work on gait training, safety, and strengthening.

Occupational therapy: Will work on ADLs.

Advance directives: Patient names [*] as the healthcare proxy. [He/She] confirmed [his/her] prior stated desires for [FULL/DNR] status. The patient has had [*] hospitalizations over the last 6 months. The patient elects [routine medical care/comfort only care] and [would/would not] desire future hospital transfers. Patient is aware of the diagnosis, condition, prognosis, and treatment plan.

[ ] Unable to reach family member/responsible party at time of admission history and physical.

[ ] able to reach family member/responsible party at time of admission history and physical.

ARTICLE

Use of a DASH Food Group Score to Predict Excess Weight Gain in Adolescent Girls in the National Growth and Health Study

Jonathan P. B. Berz, MD, MSc; Martha R. Singer, MPH; Xinxin Guo, MPH; Stephen R. Daniels, MD, PhD; Lynn L. Moore, DSc

Objective: To study the effects of selected dietary patterns, particularly a DASH (Dietary Approach to Stop Hypertension) eating pattern, on body mass index (BMI) throughout adolescence.

Design: Prospective National Growth and Health Study.

Setting: Washington, DC; Cincinnati, Ohio; and Berkeley, California.

Participants: A total of 2327 girls with 10 annual visits starting at age 9 years.

Main Exposures: Individual DASH-related food groups and a modified DASH adherence score.

Main Outcome Measure: The BMI value from measured yearly height and weight over 10 years.

Results: Longitudinal mixed modeling methods were used to assess the effects of individual DASH food groups and a DASH adherence score on BMI during 10 years of follow-up, adjusting for race, height, socioeconomic status, television viewing and video game playing hours, physical activity level, and total energy intake. Girls in the highest vs lowest quintile of the DASH score had an adjusted mean BMI of 24.4 vs 26.3 (calculated as weight in kilograms divided by height in meters squared) ($P<.05$). The strongest individual food group predictors of BMI were total fruit (mean BMI, 26.0 vs 23.6 for <1 vs ≥ 2 servings per day; $P<.001$) and low-fat dairy (mean BMI, 25.7 vs 23.2 for <1 vs ≥ 2 servings per day; $P<.001$). Whole grain consumption was more weakly but beneficially associated with BMI.

Conclusions: Adolescent girls whose diet more closely resembled the DASH eating pattern had smaller gains in BMI over 10 years. Such an eating pattern may help prevent excess weight gain during adolescence.

Arch Pediatr Adolesc Med. 2011;165(6):540-546

Author Affiliations: Sections of General Internal Medicine (Dr Berz) and Preventive Medicine and Epidemiology (Ms Singer and Dr Moore), Boston University Medical Center, Boston, and Outcome Sciences Inc, Cambridge (Ms Guo), Massachusetts; and Department of Pediatrics, University of Colorado School of Medicine, Anschutz Medical Campus, Denver (Dr Daniels).

OBESITY IS A MAJOR PUBLIC health problem, with 17% of American children overweight and 67% of adults either overweight or obese.[1-3] Excess weight during childhood leads to numerous health problems and is even associated with premature death as an adult.[4,5] Few studies have examined the relation of food-based dietary patterns with weight gain, especially in children.

The examination of food-based dietary patterns acknowledges the synergistic effects on health that food nutrients may have when eaten together.[6] One example is the DASH (Dietary Approach to Stop Hypertension) diet pattern. It emphasizes increased intakes of low-fat dairy products; fish, chicken, and lean meats (to decrease saturated fat and increase calcium levels); and nuts, fruits, whole grains, vegetables, and legumes (to increase potassium, magnesium, and dietary fiber levels).[7] The DASH eating pattern was originally studied in clinical trials of adults as a treatment for hypertension[8]; these clinical trials assessed the effects of increased fruit and vegetable intake, with or without increasing the intake of low-fat dairy products. In these studies,[9-11] the combined diet (rich in fruit, vegetables, and low-fat dairy products) led to the greatest reductions in blood pressure. The DASH pattern has also been studied in relation to the metabolic syndrome and selected cardiovascular end points.[8-14] Little has been done, however, to examine the effects of a DASH eating pattern on measures of excess weight, frequently a precursor of the aforementioned conditions. In addition, the DASH eating pattern has

See also pages 567 and 580

been infrequently studied in children, and the American Academy of Pediatrics states that there is no reason to suggest that using DASH would not be safe as long as pro-

tein and calories are consumed in quantities adequate to support child and adolescent growth and development needs.[15] In this study, we examined the effects of adherence to a DASH-style eating plan and its components on the change in body mass index (BMI) in a racially diverse sample of adolescent girls.

METHODS

STUDY POPULATION

The National Growth and Health Study was initiated by the National Heart, Lung, and Blood Institute to investigate racial differences in dietary, physical activity, family, and psychosocial factors associated with the development of obesity in black and white girls. The National Growth and Health Study enrolled 2379 girls aged 9 and 10 years in 3 cities (Washington, DC; Berkeley, California; and Cincinnati, Ohio) in 1987-1988 and observed them for 9 years. Data were collected in a longitudinal manner on 10 occasions via follow-up at annual examinations. Height and weight were measured by trained study staff using standardized assessment protocols at each examination. Additional details of the methods used for ascertaining data are described elsewhere.[16] Almost 90% of the girls originally enrolled were observed through study year 10. This study was approved by the Boston University institutional review board.

MAIN OUTCOME VARIABLE

The main outcome of interest was BMI (calculated as weight in kilograms divided by height in meters squared) at each age from 9 to 19 years.

DIETARY EXPOSURE VARIABLES

Dietary data were collected using 3-day diet records; the collection included 2 weekdays and 1 weekend day during each of 8 examination years. Participants were trained by a study nutritionist to record detailed dietary information using standard household measuring instruments for the estimation of portion sizes. Standardized debriefing was performed, and diet records that were considered unreliable by the research dietitian were excluded.

Dietary data were entered into the Nutrition Data System of the University of Minnesota, Minneapolis, to estimate the intake of total calories, macronutrients, and micronutrients.[17] The Nutrition Data System also outputs food codes for each food and each ingredient from composite foods (eg, from lasagna, macaroni and cheese, and even condiments). The Nutrition Data System food code data were combined with the US Department of Agriculture's survey food code database, the Food and Nutrient Database for Dietary Studies, version 2.0.[18] By matching these food codes, the child's average daily intake was derived in each of the 5 major food groups and in all the subgroups as defined by *Nutrition and Your Health: Dietary Guidelines for Americans* by the US Department of Agriculture.[19] Thus, we derived total servings for each group and subgroup. For example, fruit servings included fruit from all sources, such as whole fruit, fruit-based desserts, 100% fruit juice, and that portion of fruit drinks derived from fruit juice.

DASH FOOD GROUP SCORE

We created a modified DASH food group score based on a previous publication by Levitan et al.[20] This original score was designed to reflect adherence to a DASH eating pattern as described in the 2005 *Dietary Guidelines for Americans*.[21] Levitan et al[20] compared DASH scores for this scale with those of another DASH score by Fung et al[13] and found them to be moderately well correlated (r=0.61). Because the *Dietary Guidelines for Americans* differs across levels of energy intake, we used energy-specific standards for intake in each food group. The score contained 10 food groups or subgroups, 3 of which were excluded from the modified score: added sugars, discretionary fats and oils, and alcohol. Added sugars were excluded because the high intake of sugar in this population led to almost all the participants having a score of zero for this component. Discretionary fats and oils contributed nothing to the prediction of BMI in this analysis, and the alcohol component was excluded because of the ages of the girls. Therefore, we focused on the 7 DASH-related food groups in these analyses: fruits, vegetables, low-fat dairy products, total and whole grains, lean meats, and nuts, seeds, and legumes.

Low-fat dairy products were defined as those containing 2% fat or less. Lean meat was defined as fish, eggs, beef, and poultry that was 85% lean or greater. To obtain more stable estimates of intake, we included only girls with 2 or more sets of 3-day diet records collected between ages 9 and 17 years (2330 of the original 2379 participants). One girl with an average intake of less than 1000 kcal/d and 1 with an average of more than 3500 kcal/d at ages 9 to 17 years were excluded from the study, as well as 1 girl with absent physical activity data, leaving a final sample of 2327 girls with available data who were included in these analyses.

We followed the scoring protocol of Levitan et al.[20] Each food group was assigned a score ranging from 0 to 1. For total grains, meats, low-fat dairy products, and nuts, seeds, and legumes, participants with intake meeting the guidelines were assigned 1 point. Those with intake above the recommended levels were assigned partial points as follows: 1 minus the percentage of intake over the guideline. For intake below the guideline level, partial points were assigned by dividing the actual intake by the recommended intake. For fruits, vegetables, and whole grains where optimal intake was deemed to be at or greater than recommended, a full point was assigned to those who consumed the recommended level of intake or higher. Partial points were given only for those who had less than the recommended intake. Because DASH recommends that most grains be whole, we used 50% of the total grain recommended as the goal for whole grain intake in accord with American Heart Association guidelines.[15] The total DASH score was, thus, a sum of the scores for each individual food group.

POTENTIAL CONFOUNDING VARIABLES

Potential confounding factors that were evaluated for inclusion in these analyses included race, height at each age, socioeconomic status (SES), physical activity level, television viewing and video game playing (hours per day), total energy intake, and other dietary factors.[22] The SES was classified as low, moderate, or high by combining information about parental income and education. Low-SES families were those with incomes of less than $10 000 per year or parental education level of less than high school; high-SES families included those with incomes of at least $40 000 per year and at least a high school education. All the other participants were classified as moderate SES. Physical activity was measured at each visit using a validated questionnaire that asked the participants to report the frequency and duration per week (during the school year and summer) of participation in a variety of structured physical activities in the past year.[23] These data were combined with published information on metabolic equivalent levels to obtain a final score estimating daily energy expenditure.[24-26]

Table 1. Characteristics of the Study Population by Quintile of DASH Adherence Score[a]

Characteristic	Quintile 1 (n=465)	2 (n=465)	3 (n=466)	4 (n=466)	5 (n=465)
DASH adherence score, range[b]	1.3-2.6	2.6-2.9	2.9-3.3	3.3-3.6	3.6-5.2
Food group DASH score, mean (SD), servings/d[c]					
Total grains	5.74 (1.72)	6.16 (1.53)	6.40 (1.48)	6.45 (1.41)	6.48 (1.24)
Vegetables	1.63 (0.65)	2.07 (0.85)	2.18 (0.82)	2.24 (0.88)	2.38 (0.86)
Lean meats	1.34 (0.71)	1.67 (0.82)	1.75 (0.91)	2.05 (1.00)	2.13 (1.07)
Fruits	0.80 (0.53)	0.98 (0.59)	1.16 (0.69)	1.43 (0.80)	1.93 (1.05)
Low-fat dairy products	0.63 (0.40)	0.78 (0.51)	0.97 (0.57)	1.12 (0.67)	1.52 (0.76)
Whole grains	0.36 (0.28)	0.43 (0.31)	0.51 (0.32)	0.59 (0.36)	0.77 (0.43)
Nuts/seeds/legumes	0.18 (0.23)	0.24 (0.24)	0.27 (0.20)	0.32 (0.20)	0.38 (0.20)
Physical activity score, mean (SD)	18.1 (9.6)	18.7 (10.1)	19.9 (10.0)	20.5 (10.7)	23.2 (10.5)
Television viewing and video game playing, mean (SD), h/d	5.3 (2.1)	5.4 (2.1)	5.2 (2.1)	4.9 (2.2)	3.8 (2.1)
Total energy intake, mean (SD), kcal/d	1686 (388)	1872 (369)	1912 (374)	1949 (354)	1944 (290)
Race, row %					
White (n = 1139)	14.8	15.3	19.2	20.6	30.0
Black (n = 1189)	24.9	24.5	20.8	19.4	10.4
SES, row %					
Low (n = 548)	26.9	27.2	21.0	16.3	8.6
Middle (n = 996)	21.4	20.1	22.3	19.4	16.9
High (n = 784)	13.4	14.8	16.5	23.5	31.9

Abbreviations: DASH, Dietary Approach to Stop Hypertension; SES, socioeconomic status.
[a] $P < .001$ for all.
[b] The maximum possible DASH score was 7.
[c] Dietary intakes are averages across approximately 20 days of diet records.

STATISTICAL ANALYSIS

Confounders were evaluated using a forward selection method. Factors determined to be confounders and those that were independent predictors of the outcome were retained in the final models (ie, race, height, SES, physical activity level, television viewing and video game playing, and total energy).

Categories of average intake in each of the DASH food groups were determined by balancing information about the distributions of the intake population (which affects study power) with the recommended intake levels. For example, DASH recommended intake level of fruits is 4 to 5 servings per day, but the category cutoff values used in the analysis were less than 1, 1 to less than 2, and 2 or more servings per day because few participants actually consumed 4 to 5 servings. Additional analyses were conducted to evaluate the sensitivity of the results to subtle changes in category definitions.

To determine the association between level of DASH food group intake and BMI over time, we used longitudinal data analysis methods, accounting for correlated observations from the repeated-measures data. In separate models, each categorical food group variable was entered as the main exposure variable, with age as an interaction factor, while controlling for fixed and changing potential confounders as previously described. This allowed us to estimate the adjusted mean BMI at each age in each category of intake for each food group. Analyses were conducted using an unstructured covariance matrix in Proc Mixed with the repeated option in SAS (SAS Institute Inc, Cary, North Carolina).[22] The same longitudinal mixed modeling methods were used to estimate the adjusted mean BMI at each age according to quintile of DASH score.

In each model, the interaction of age and food group was examined first. When a significant interaction was found, further testing was performed to evaluate differences between the slopes, intercepts, and BMIs at the end of follow-up. When there was no significant age–food group interaction ($P > .05$), the statistical significance of the fixed effects for the main dietary exposure variable was examined (using type III sums of squares). Approximate linearity of the relationship between age and BMI was assumed. All analyses were performed using a commercially available statistical software program (SAS, version 9.1).

RESULTS

STUDY CHARACTERISTICS

Dietary and population characteristics by quintile of DASH score are presented in **Table 1**. The overall mean DASH adherence score was 3.1, with a median of 3.1 (range, 1.3-5.2). Food group means in each quintile show that higher DASH scores were associated with higher intake in most food groups. Higher DASH scores were also associated with higher total energy intake. Black participants and those with lower SES were more likely to be in a lower quintile of DASH scores. In addition, there was higher mean physical activity and lower mean television viewing and video game playing hours in the highest quintile of the DASH score.

The distribution of actual intakes for each food component of the DASH score is given in **Table 2**, along with the recommended DASH intakes. Even study participants in the 95th percentile of intake had relatively low intake of fruits, vegetables, whole grains, and low-fat dairy products compared with the DASH recommendations. The average intake of added sugars was approximately 10 times higher than recommended. Discretionary fat intake was also relatively high, although no direct equivalent DASH recommendation applies.

Table 2. Actual and Recommended Intakes of DASH Food Groups[a]

Food Group	Servings per Day, Percentile[b]			DASH Recommendations, Servings per Day
	5th	50th	95th	
Discretionary fat	7.5	11.7	17.7	NA
Grains	4.0	6.2	8.9	6
Added sugar	3.7	7.2	12.0	0.7
Vegetables	1.0	2.0	3.6	3-4
Lean meat	0.5	1.6	3.6	1-2
Fruits	0.3	1.1	2.9	4
Whole grains	0.1	0.5	1.2	3
Nuts/seeds/legumes	0	0.2	0.7	0.5

Abbreviations: DASH, Dietary Approach to Stop Hypertension; NA, no DASH recommendation applies.
[a] Based on intake of 1600 kcal/d.
[b] Serving sizes for each food group: lean meat, 3 oz; low-fat dairy, 1 cup milk or yogurt or 1.5 oz cheese; nuts/seeds/legumes, ⅓ cup nuts, 2 tbsp seeds, and ½ cup cooked dry beans; (whole) grains, 1 slice of bread, 1 oz dry cereal, and ½ cup cooked rice, pasta, or cereal; vegetables, 1 cup raw/leafy, ½ cup cooked, and 6 oz vegetable juice; fruits, 6 oz fruit juice, 1 medium piece, ¼ cup dried, and ½ cup fresh, frozen, or canned; added sugar, 1 tbsp; and discretionary fat, 5 g/1 tsp.

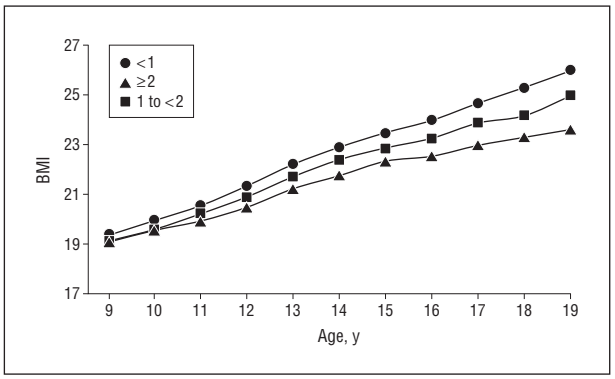


Figure 1. Body mass index (BMI) (calculated as weight in kilograms divided by height in meters squared) over 10 years associated with fruit consumption (mean servings per day). Adjusted for race, height, socioeconomic status, physical activity level, television viewing and video game playing hours per day, added sugar, total dairy, vegetables, total grains, nuts/seeds/legumes, processed and nonprocessed meat, and total energy intake. Slopes: $P<.001$ for differences between each line. Difference in BMI at end of follow-up: less than 1 vs 1 to less than 2 servings, less than 1 vs 2 to less than 8 servings, and 1 to less than 2 vs 2 to less than 8 servings, $P<.001$ for all.

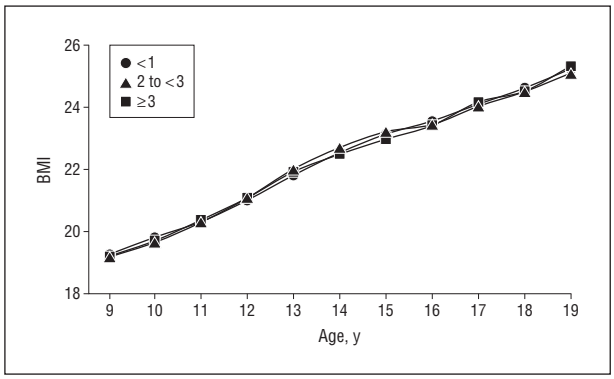


Figure 2. Body mass index (BMI) (calculated as weight in kilograms divided by height in meters squared) over 10 years associated with vegetable consumption (mean servings per day). Adjusted for race, height, socioeconomic status, physical activity level, television viewing and video game playing hours per day, added sugar, total dairy, fruit, total grains, nuts/seeds/legumes, processed and nonprocessed meat, and total energy intake. For the overall difference between age and total grain interaction, $P=.16$. For the overall difference between intake categories, $P>.99$. The BMI at the end of follow-up did not statistically differ significantly among the 3 consumption groups.

INTAKE OF INDIVIDUAL FOOD GROUPS VS BMI

Figures 1, **2**, **3**, and **4** show the adjusted mean BMI at each age associated with average intake in 4 DASH food groups: fruits, vegetables, whole grains, and low-fat dairy products. Participants who consumed 2 or more servings of fruit per day had the smallest gain in BMI over time ($P<.001$) and the lowest BMI at the end of follow-up (23.6, 25.0, and 26.0 for low, moderate, and higher intakes of fruit, respectively) (Figure 1 and **Table 3**). No differences were noted in BMI according to intake of vegetables (Figure 2 and Table 3). Highest (vs lowest) whole grain intake conferred lower BMI increases over time ($P=.01$) and a lower BMI at the end of follow-up (Figure 3 and Table 3). Higher intake of low-fat dairy products led to lower BMI gains over time (Figure 4 and Table 3). In data not shown, we compared models including and excluding total energy and total and saturated fat as a percentage of energy intake. No substantive differences in the BMI trajectory were observed.

DASH FOOD GROUP SCORE AS A PREDICTOR OF BMI

Figure 5 shows adjusted mean BMIs at each age associated with quintiles of the DASH score, averaged over ages 9 to 17 years. Girls in the highest quintile had the smallest gains in BMI over time and the lowest BMIs at the end of follow-up (Table 3). In addition, at age 19 years, those in the lowest quintile of the DASH score (compared with those in the highest quintile) had a mean BMI that was greater than the threshold for overweight as defined by the 85th percentile for age.[27]

COMMENT

In this longitudinal cohort of adolescent girls, we found that higher adherence to a DASH-style diet resulted in a consistently lower BMI between the ages of 9 and 19 years. These findings were stable over a 10-year follow-up and

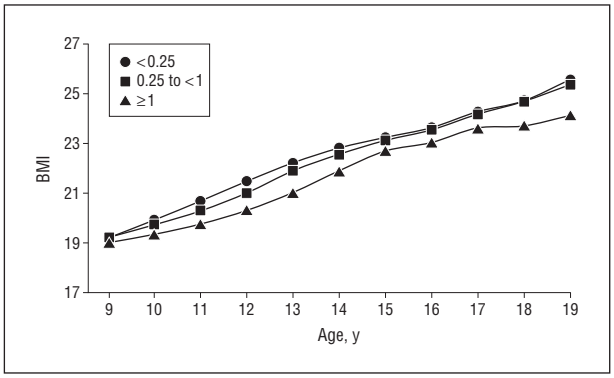


Figure 3. Body mass index (BMI) (calculated as weight in kilograms divided by height in meters squared) over 10 years associated with whole grain intake (mean servings per day). Adjusted for race, height, socioeconomic status, physical activity level, television viewing and video game playing hours per day, added sugar, total dairy, fruit, vegetables, nuts/seeds/legumes, processed and nonprocessed meat, and total energy intake. Slopes: less than 0.25 vs 0.25 to less than 1 serving, $P=.62$; less than 0.25 vs 1 to less than 7 servings, $P=.05$; and 0.25 to less than 1 vs 1 to less than 7 servings, $P=.01$. The BMI at the end of follow-up: less than 0.25 vs 0.25 to less than 1 serving, $P=.51$; and less than 0.25 vs 1 to less than 7 servings, $P=.002$.

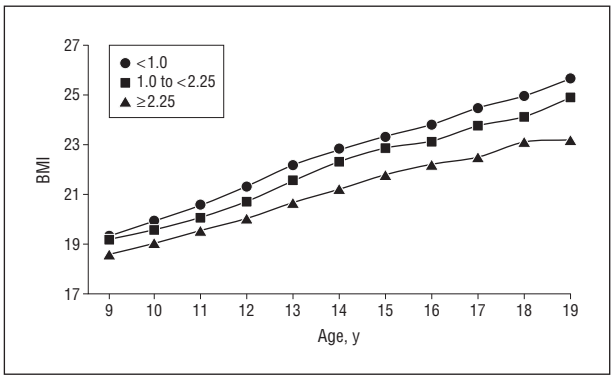


Figure 4. Body mass index (BMI) (calculated as weight in kilograms divided by height in meters squared) over 10 years associated with low-fat dairy products intake (mean servings per day). Adjusted for race, height, socioeconomic status, physical activity level, television viewing and video game playing hours per day, added sugar, fruit, vegetables, total grains, nuts/seeds/legumes, processed and nonprocessed meat, and total energy intake. Slopes: less than 1 vs 1 to less than 2.25 servings, $P=.001$; less than 1 vs 2.25 to less than 5 servings, $P<.001$; and 1 to less than 2.25 vs 2.25 to less than 5 servings, $P=.02$. The BMI at the end of follow-up: less than 1 vs 1 to less than 2.25 servings, $P=.007$; and less than 1 vs 2.25 to less than 5 servings, $P<.001$.

after controlling for nondietary factors associated with eating patterns and excess weight gain.

Few studies have examined dietary patterns in children or used longitudinal data to examine their effects on weight gain. A cross-sectional study[28] of Korean preschool children found that a diet pattern that shares components of the DASH eating pattern was not associated with measured weight status. One longitudinal study[29] of women showed that a pattern of intake lower in fruit, vegetables, and low-fat foods was associated with a greater chance of becoming overweight, and another study[30] showed that a diet consisting of many components present in the DASH pattern resulted in smaller gains in BMI over time.

The present findings for the DASH score were mirrored by the effects of some of the individual food group components. In particular, higher consumption of fruits, whole grains, and low-fat dairy products led to less weight

Table 3. BMI at Baseline and End of Follow-up for 4 DASH Food Groups and DASH Adherence Score[a]

Food Group and Intake, Mean Servings per Day	Participants, No.	BMI, Mean (SD)	
		Baseline	End of Follow-up
Fruits			
<1	1060	19.4 (0.16)	26.0 (0.19)
1 to <2	882	19.1 (0.16)	25.0 (0.21)
≥2	385	19.1 (0.22)	23.6 (0.32)
Vegetables			
<2	1192	19.3 (0.16)	25.2 (0.19)
2 to <3	833	19.2 (0.17)	25.3 (0.22)
≥3	302	19.2 (0.27)	25.1 (0.36)
Whole grains			
<0.25	531	19.2 (0.20)	25.5 (0.26)
0.25 to <1	1543	19.2 (0.14)	25.3 (0.16)
≥1	253	19.0 (0.26)	24.1 (0.38)
Low-fat dairy products			
<1.0	1363	19.3 (0.15)	25.7 (0.17)
1.0 to <2.25	834	19.2 (0.17)	24.9 (0.22)
≥2.25	130	18.6 (0.37)	23.2 (0.55)
DASH adherence score			
Quintile 1	465	19.7 (0.21)	26.3 (0.28)
Quintile 2	465	19.0 (0.20)	24.9 (0.28)
Quintile 3	466	19.3 (0.20)	25.2 (0.28)
Quintile 4	466	19.0 (0.20)	25.3 (0.29)
Quintile 5	465	19.1 (0.20)	24.4 (0.30)

Abbreviations: BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); DASH, Dietary Approach to Stop Hypertension.

[a]The mixed models did not drop the girls below 1000 kcal/d (1 girl) and above 3500 kcal/d (1 girl). Dropping them and rerunning the program gives results that only differ at the second decimal place, so the data do not need to be changed. However, data are from 2327 instead of 2328 girls because one girl is missing activity at all ages.

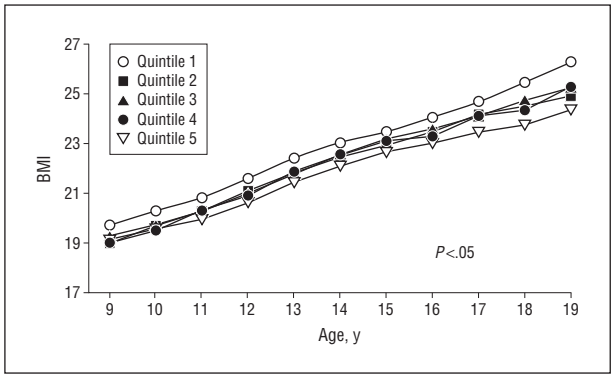


Figure 5. Body mass index (BMI) (calculated as weight in kilograms divided by height in meters squared) over 10 years associated with DASH (Dietary Approach to Stop Hypertension) adherence score (quintile). Adjusted for race, height, socioeconomic status, physical activity level, television viewing and video game playing hours per day, and total energy intake. Slopes: quintiles 1 to 4 vs quintile 5, $P<.05$. The BMI at the end of follow-up: quintiles 1, 3, and 4 vs quintile 5, $P<.05$.

gain. The observed fruit intake in this study was well below the DASH recommendation of 4 servings per day; on average, at 9 to 17 years of age, only 15% of girls reached this goal. In addition, higher vegetable consumption was not associated with decreased weight gain over time. It is possible that relatively low intakes of vegetables and the narrow range of types of vegetables con-

sumed (ie, a predominance of starchy vegetables) may explain the absence of a protective effect. Indeed, in a subanalysis, a higher intake of nonstarchy vegetables was associated with a lower BMI at the end of follow-up (data not shown).

The diet records used in the present study may provide more precise ascertainment of total fruit intake in children than do Frequency Food Questionnaire methods because we extracted fruit servings from composite dishes, as previously described. Thus, the present study is likely to have less nondifferential misclassification of diet exposures and a greater ability to detect meaningful associations between diet and BMI.

Higher low-fat dairy product consumption resulted in smaller increases in BMI during adolescence. Data on dairy consumption and excess weight change during adolescence show mixed results in the larger literature. Two small studies[31,32] found no effect of dairy intake on weight gain, and another study[33] showed an adverse effect of higher dairy consumption on weight gain, even for low-fat milk, although the effect seemed to be mediated by excess energy intake. In contrast, 2 other studies,[34,35] one using diet records and another using the Frequency Food Questionnaire, found that higher dairy intake protected young adults from excess weight gain.

The present study may be the largest long-term study using diet records that has shown a protective effect of dairy intake on weight gain. Dairy may act to decrease weight gain through a variety of possible mechanisms, including an association with a healthier diet in general; its protein content has been shown to increase satiety.[36]

Higher intakes of whole grains were associated with decreased BMI gains during study follow-up. Although there are few studies on grain intake and weight in children, the present finding is in line with other studies[37-40] showing grain to be protective. Although the level of whole grain intake across this study population was low and well below the target threshold of 50% of total grains, it was nonetheless protective. Whole grain intake may result in less weight gain via its higher fiber content and, thus, higher satiety or as a marker for a healthier lifestyle, something we may not have been able to completely capture in the multivariate models.

The study strengths include use of a large socioeconomically and geographically diverse sample that incorporates more than 50% black girls, a population particularly beset by the obesity epidemic. An additional strength is the availability of detailed dietary information that allows us to examine the change in BMI in relation to food group exposure, a method that has seen little study in the adolescent literature so far. Finally, the use of repeated measures collected in a longitudinal manner over 9 years of study increased power substantially and likely decreased random variation in exposure and covariate data.

A limitation of this study is the low level of intake in certain food groups that may have restricted our ability to detect true beneficial effects of these food groups. In addition, there are food groups that other studies have found to be important predictors of obesity, such as sugar-sweetened beverages that are not a part of the DASH eating pattern and unavailable in this analysis, that may be important predictors of excess weight gain.[41,42] Finally, it is possible that there is biased reporting for some dietary factors, especially in obese individuals, that could affect the results of these analyses. However, the longitudinal nature of the study and the multiple measures of dietary intake beginning in preadolescence suggest that this explanation for these findings is unlikely. In conclusion, greater consistency with the DASH eating plan resulted in lower excess weight gains in girls from early adolescence to young adulthood.

Accepted for Publication: December 23, 2010.
Correspondence: Jonathan P. B. Berz, MD, MSc, Section of General Internal Medicine, Boston University Medical Center, 801 Massachusetts Ave, Second Floor, Boston, MA 02118 (Jonathan.Berz@bmc.org).
Author Contributions: Drs Berz and Moore had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Berz and Moore. *Acquisition of data:* Singer, Daniels, and Moore. *Analysis and interpretation of data:* Berz, Guo, and Moore. *Drafting of the manuscript:* Berz and Guo. *Critical revision of the manuscript for important intellectual content:* Berz, Singer, Daniels, and Moore. *Statistical analysis:* Guo and Moore. *Obtained funding:* Daniels and Moore. *Administrative, technical, and material support:* Singer and Moore. *Study supervision:* Moore.
Financial Disclosure: None reported.
Funding/Support: This work was supported by grant 5R21DK075068 from the National Institute of Diabetes and Digestive and Kidney Diseases.
Role of the Sponsor: No sponsors or funders had any role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; and in the preparation, review, or approval of the manuscript.
Additional Contributions: Di Gao, AS, Caroline Apovian, MD, Gheorghe Doros, PhD, and M. Loring Bradlee, MS, assisted with statistical analysis, data presentation preparation, and insight.

REFERENCES

1. Chang VW, Lauderdale DS. Income disparities in body mass index and obesity in the United States, 1971-2002. *Arch Intern Med*. 2005;165(18):2122-2128.
2. Hedley AA, Ogden CL, Johnson CL, Carroll MD, Curtin LR, Flegal KM. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999-2002. *JAMA*. 2004;291(23):2847-2850.
3. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. *JAMA*. 2006;295(13):1549-1555.
4. van Dam RM, Willett WC, Manson JE, Hu FB. The relationship between overweight in adolescence and premature death in women. *Ann Intern Med*. 2006;145(2):91-97.
5. Daniels SR, Arnett DK, Eckel RH, et al. Overweight in children and adolescents: pathophysiology, consequences, prevention, and treatment. *Circulation*. 2005;111(15):1999-2012.
6. Jacobs DR Jr, Tapsell LC. Food, not nutrients, is the fundamental unit in nutrition. *Nutr Rev*. 2007;65(10):439-450.
7. Sacks FM, Obarzanek E, Windhauser MM, et al. Rationale and design of the Dietary Approaches to Stop Hypertension trial (DASH). A multicenter controlled-feeding study of dietary patterns to lower blood pressure. *Ann Epidemiol*. 1995;5(2):108-118.
8. Appel LJ, Moore TJ, Obarzanek E, et al; DASH Collaborative Research Group. A clinical trial of the effects of dietary patterns on blood pressure. *N Engl J Med*. 1997;336(16):1117-1124.

9. Champagne CM. Dietary interventions on blood pressure: the Dietary Approaches to Stop Hypertension (DASH) trials. *Nutr Rev*. 2006;64(2, pt 2):S53-S56.
10. Moore LL, Singer MR, Bradlee ML, et al. Intake of fruits, vegetables, and dairy products in early childhood and subsequent blood pressure change. *Epidemiology*. 2005;16(1):4-11.
11. Moore TJ, Conlin PR, Ard J, Svetkey LP. DASH (Dietary Approaches to Stop Hypertension) diet is effective treatment for stage 1 isolated systolic hypertension. *Hypertension*. 2001;38(2):155-158.
12. Folsom AR, Parker ED, Harnack LJ. Degree of concordance with DASH diet guidelines and incidence of hypertension and fatal cardiovascular disease. *Am J Hypertens*. 2007;20(3):225-232.
13. Fung TT, Chiuve SE, McCullough ML, Rexrode KM, Logroscino G, Hu FB. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. *Arch Intern Med*. 2008;168(7):713-720.
14. Mellen PB, Gao SK, Vitolins MZ, Goff DC Jr. Deteriorating dietary habits among adults with hypertension: DASH dietary accordance, NHANES 1988-1994 and 1999-2004. *Arch Intern Med*. 2008;168(3):308-314.
15. Gidding SS, Dennison BA, Birch LL, et al; American Heart Association. Dietary recommendations for children and adolescents: a guide for practitioners [published correction appears in *Pediatrics*. 2006;118(3):1323]. *Pediatrics*. 2006;117(2):544-559.
16. Obesity and cardiovascular disease risk factors in black and white girls: the NHLBI Growth and Health Study. *Am J Public Health*. 1992;82(12):1613-1620.
17. Schakel SF, Sievert YA, Buzzard IM. Sources of data for developing and maintaining a nutrient database. *J Am Diet Assoc*. 1988;88(10):1268-1271.
18. Agricultural Research Service. *USDA Food and Nutrient Database for Dietary Studies, 2.0*. Beltsville, MD: Agricultural Research Service, Food Surveys Research Group; 2006.
19. *Nutrition and Your Health: Dietary Guidelines for Americans*. 5th ed. Washington, DC: US Government Printing Office; 2000.
20. Levitan EB, Wolk A, Mittleman MA. Consistency with the DASH diet and incidence of heart failure. *Arch Intern Med*. 2009;169(9):851-857.
21. *Dietary Guidelines for Americans*. 2005. http://www.health.gov/dietaryguidelines/dga2005/document/pdf/dga2005.pdf. Accessed October 17, 2008.
22. Littell RC, Pendergast J, Natarajan R. Modelling covariance structure in the analysis of repeated measures data. *Stat Med*. 2000;19(13):1793-1819.
23. Kimm SY, Glynn NW, Kriska AM, et al. Longitudinal changes in physical activity in a biracial cohort during adolescence. *Med Sci Sports Exerc*. 2000;32(8):1445-1454.
24. Harrell JS, McMurray RG, Baggett CD, Pennell ML, Pearce PF, Bangdiwala SI. Energy costs of physical activities in children and adolescents. *Med Sci Sports Exerc*. 2005;37(2):329-336.
25. Ridley K, Olds TS. Assigning energy costs to activities in children: a review and synthesis. *Med Sci Sports Exerc*. 2008;40(8):1439-1446.
26. Ridley K, Ainsworth BE, Olds TS. Development of a compendium of energy expenditures for youth. *Int J Behav Nutr Phys Act*. 2008;5:45.
27. Statistics National Center for Health Statistics. BMI for age charts: CDC growth charts: United States. May 30, 2000. http://www.cdc.gov/nchs/data/series/sr_11/sr11_246.pdf. Accessed June 1, 2009.
28. Shin KO, Oh SY, Park HS. Empirically derived major dietary patterns and their associations with overweight in Korean preschool children. *Br J Nutr*. 2007;98(2):416-421.
29. Quatromoni PA, Copenhafer DL, D'Agostino RB, Millen BE. Dietary patterns predict the development of overweight in women: the Framingham Nutrition Studies. *J Am Diet Assoc*. 2002;102(9):1239-1246.
30. Newby PK, Muller D, Hallfrisch J, Qiao N, Andres R, Tucker KL. Dietary patterns and changes in body mass index and waist circumference in adults. *Am J Clin Nutr*. 2003;77(6):1417-1425.
31. Phillips SM, Bandini LG, Cyr H, Colclough-Douglas S, Naumova E, Must A. Dairy food consumption and body weight and fatness studied longitudinally over the adolescent period. *Int J Obes Relat Metab Disord*. 2003;27(9):1106-1113.
32. Dixon LB, Pellizzon MA, Jawad AF, Tershakovec AM. Calcium and dairy intake and measures of obesity in hyper- and normocholesterolemic children. *Obes Res*. 2005;13(10):1727-1738.
33. Berkey CS, Rockett HR, Willett WC, Colditz GA. Milk, dairy fat, dietary calcium, and weight gain: a longitudinal study of adolescents. *Arch Pediatr Adolesc Med*. 2005;159(6):543-550.
34. Moore LL, Bradlee ML, Gao D, Singer MR. Low dairy intake in early childhood predicts excess body fat gain. *Obesity (Silver Spring)*. 2006;14(6):1010-1018.
35. Pereira MA, Jacobs DR Jr, Van Horn L, Slattery ML, Kartashov AI, Ludwig DS. Dairy consumption, obesity, and the insulin resistance syndrome in young adults: the CARDIA Study. *JAMA*. 2002;287(16):2081-2089.
36. Luhovyy BL, Akhavan T, Anderson GH. Whey proteins in the regulation of food intake and satiety. *J Am Coll Nutr*. 2007;26(6):704S-712S.
37. Liu S, Willett WC, Manson JE, Hu FB, Rosner B, Colditz G. Relation between changes in intakes of dietary fiber and grain products and changes in weight and development of obesity among middle-aged women. *Am J Clin Nutr*. 2003;78(5):920-927.
38. Lutsey PL, Jacobs DR Jr, Kori S, et al. Whole grain intake and its cross-sectional association with obesity, insulin resistance, inflammation, diabetes and subclinical CVD: the MESA Study. *Br J Nutr*. 2007;98(2):397-405.
39. Newby PK, Peterson KE, Berkey CS, Leppert J, Willett WC, Colditz GA. Dietary composition and weight change among low-income preschool children. *Arch Pediatr Adolesc Med*. 2003;157(8):759-764.
40. McKeown NM, Meigs JB, Liu S, Wilson PWF, Jacques PF. Whole-grain intake is favorably associated with metabolic risk factors for type 2 diabetes and cardiovascular disease in the Framingham Offspring Study. *Am J Clin Nutr*. 2002;76(2):390-398.
41. Ludwig DS, Peterson KE, Gortmaker SL. Relation between consumption of sugar-sweetened drinks and childhood obesity: a prospective, observational analysis. *Lancet*. 2001;357(9255):505-508.
42. Francis LA, Lee Y, Birch LL. Parental weight status and girls' television viewing, snacking, and body mass indexes. *Obes Res*. 2003;11(1):143-151.

Education is the ability to meet life's situations.

—Dr John Hibbon, former president of Princeton

Hindawi Publishing Corporation
Journal of Environmental and Public Health
Volume 2011, Article ID 435078, 12 pages
doi:10.1155/2011/435078

Research Article

The Use of Mixed Models for the Analysis of Mediated Data with Time-Dependent Predictors

Emily A. Blood[1] and Debbie M. Cheng[2]

[1] *Children's Hospital Boston and Harvard Medical School, 300 Longwood Avenue, Boston, MA 02115, USA*
[2] *Boston University School of Public Health, 801 Massachusetts Avenue 3rd Floor, Boston, MA 02118, USA*

Correspondence should be addressed to Emily A. Blood, emily.blood@childrens.harvard.edu

Received 15 October 2010; Revised 14 February 2011; Accepted 24 February 2011

Academic Editor: Pam R. Factor-Litvak

Linear mixed models (LMMs) are frequently used to analyze longitudinal data. Although these models can be used to evaluate mediation, they do not directly model causal pathways. Structural equation models (SEMs) are an alternative technique that allows explicit modeling of mediation. The goal of this paper is to evaluate the performance of LMMs relative to SEMs in the analysis of mediated longitudinal data with time-dependent predictors and mediators. We simulated mediated longitudinal data from an SEM and specified delayed effects of the predictor. A variety of model specifications were assessed, and the LMMs and SEMs were evaluated with respect to bias, coverage probability, power, and Type I error. Models evaluated in the simulation were also applied to data from an observational cohort of HIV-infected individuals. We found that when carefully constructed, the LMM adequately models mediated exposure effects that change over time in the presence of mediation, even when the data arise from an SEM.

1. Introduction

In clinical research, both outcomes and predictors are frequently collected repeatedly over time and complex mediated relationships may be present among the variables of interest. For example, in a study of the relationship between alcohol use and HIV disease progression, heavy alcohol consumption may affect antiretroviral therapy (ART) adherence which, in turn, affects CD4 cell count. However, alcohol consumption itself may also directly affect CD4 count. If the goal is to evaluate the total effect of the main independent variable (e.g., alcohol consumption) on the outcome (CD4 count), a single linear mixed effects model (LMM) [1] could be fit to the data. LMMs account for correlation among repeated observations within an individual and are frequently used to analyze longitudinal data. To disentangle the direct versus indirect effects of alcohol use on HIV disease progression, however, a series of LMMs could be fit according to the steps described by Baron and Kenny [2] and demonstrated by Krull and MacKinnon [3] in the mixed model setting. In contrast, if these data were analyzed with a structural equation model (SEM) [4], variables in the causal pathway could be modeled directly by incorporating adherence into the SEM as a mediating variable between heavy alcohol consumption and HIV disease progression. Given the objective is to evaluate the total effect of the main independent variable, it is unclear whether there are benefits to modeling the mediated relationship in terms of bias, coverage, and power for the primary study aim.

Tradeoffs between the use of SEMs and LMMs have been previously evaluated in general settings, and the equivalence of LMMs and SEMs in some settings without mediation has been well documented in the SEM literature [5–12]. The potential advantages of using SEMs over LMMs to analyze longitudinal or hierarchical data include the capacity to explicitly model complex relationships such as mediation [4, 5, 7, 13–16], the flexibility in modeling covariance structures [7, 15], the availability of fit indices [8, 9], and the capability to account for measurement error [5, 9, 10, 15]. One disadvantage is the potential complexity of the SEM model and, therefore, the possibility of model misspecification. In addition, from a practical perspective, the SEM may be less convenient to implement given the need for specialized software. Nonetheless, its flexibility and capacity to directly

model variables in the causal pathway make it an appealing modeling technique for mediated longitudinal data.

In the absence of mediation, the type of SEM evaluated in this paper is often referred to as a latent growth curve model [13, 17–20]. Incorporating mediation into a latent growth curve framework has been demonstrated using either a time-invariant mediating factor that influences the latent intercept and slope factors of an outcome trajectory [14] or a time-varying mediator that assumes a parallel growth process in which both the mediator and outcome follow growth trajectories [21, 22]. For both of these approaches, mediation occurs at the random effect level (individual), rather than the observation level and, therefore, cannot vary over time. Modeling mediation that occurs at multiple levels in longitudinal data has been described using separate linear mixed effects models [3, 23, 24]. These multilevel models allow for mediation at the individual as well as observation level, but indirect and total effects are estimated from separate regressions. In the multi-level context, methods for assessing mediation at the observation level have been described with the added complexity that all mediated effects are random [25, 26]. Finally, longitudinal mediation has been described outside of the latent growth curve framework using autoregressive structural equation models [24, 27]. These models assume change over time, where the correlation between observations is not due to underlying random effects (latent intercept and latent slope), but rather results from direct association between an outcome and its value at a previous time point. Autoregressive models are, therefore, not a direct extension of LMMs but represent an alternative approach to model mediated longitudinal data. In this paper, we examine an SEM in which mediation is present at each time point and can, therefore, vary at the observation level. We do not assume that the mediator follows a parallel growth process and assume fixed, not random, effects of the mediator on the outcome. The mediated effects are estimated simultaneously rather than through separate multi-level models.

The performance of LMMs relative to SEMs in a longitudinal data setting with a predictor and mediator both measured only at baseline with longitudinal outcomes has previously been studied [28]. The LMM was accurate and efficient in a variety of settings in estimating the total effect of the main independent variable. The main advantage of the SEM was found to be the ability to simultaneously model the direct and indirect effects of the main independent variable. The objective of this study is to extend this previous work to the setting where the predictor and mediator are both time dependent with fixed effects that change across time.

2. Methods

In the current study, we evaluate the performance of the LMM relative to the SEM in the analysis of mediated longitudinal data with a time-varying predictor and mediator. As an example, we consider a prospective cohort study assessing the effect of heavy alcohol consumption on HIV disease progression [29]. The continuous outcome, CD4 cell count, is denoted by Y_j. The main independent variable, heavy alcohol consumption, is a time-varying binary variable denoted by z_j; ART adherence, the mediating variable, is a time-varying variable denoted by M_j; and baseline age, a continuous covariate, is denoted by w. ART adherence is a mediator if the primary independent variable, heavy alcohol use, affects CD4 count indirectly through ART adherence. In addition to indirect effects, heavy alcohol use may also have a direct effect on CD4 cell count that is not mediated by ART adherence or other variables. We focus on a setting where the primary aim is to determine the total effect (direct and indirect effect) of heavy alcohol use on CD4 cell count while appropriately accounting for the mediating effect of ART adherence. We arbitrarily assume there are six time points at which the outcome, predictor, and mediator are measured. Time is represented by t_j ($j = 1, 2, \ldots, 6$), and times are assumed equally spaced. In this setting, an LMM could be used to evaluate the total effect of alcohol consumption on CD4 cell count while accounting for correlation due to multiple assessments from the same individual and confounding effects of covariates. Using a LMM would not, however, allow for directly modeling mediation among the variables. SEMs are an alternative approach with the advantage of simultaneous modeling of direct and indirect effects of alcohol consumption on CD4 cell count. The objective of this paper is to evaluate whether the LMM performs adequately relative to the SEM when the goal is to determine the total effect of alcohol consumption, rather than to evaluate whether a variable (e.g., adherence) is a mediator. A series of simulation studies is carried out to evaluate the performance of several LMMs and SEMs under different conditions. We also describe the application of the various models to data from a prospective cohort study evaluating the impact of alcohol use on HIV disease progression.

2.1. General SEM Incorporating Mediation. There are two components to an SEM, the measurement model and the structural model [4]. The measurement model relates unobserved latent variables and covariates to outcomes and exposure indicators. In the measurement model, outcomes are observed variables, while predictors may be observed or latent variables. This model attempts to capture measurement error in observed variables. In the SEM, the repeated observations of CD4 count are the outcomes in the measurement model. The predictors in this model include underlying individual intercept and slope variables as well as time-varying primary independent variable (heavy alcohol use) and the time-varying mediator (ART adherence).

The second component to an SEM, the structural model, models latent variables as a function of observed variables and other latent variables. This model attempts to capture individual variation in the latent variables. In our model, the underlying individual intercept and slope variables are treated as latent variables and modeled as the outcomes of the structural model. In the case of the SEM incorporating time-varying mediators, the repeated mediators (ART adherence), while not latent variables, are also outcomes predicted with

some error by the time-varying primary independent variable (alcohol use) so they are incorporated in the structural model.

The general SEM incorporating mediation is described in the following equations. The subject index has been dropped in the equations below for simplicity:

$$Y_j = U_1 + t_j U_2 + \lambda_j M_j + \kappa_j z_j + \epsilon_j, \quad (1)$$

$$U_1 = \alpha_1 + \gamma_2 w + \zeta_1, \quad (2)$$

$$U_2 = \alpha_2 + \zeta_2. \quad (3)$$

for $j =$ 1 to 6,

$$M_j = \alpha_3 + \gamma_{1j}\, z_j + \zeta_{2+j}, \quad (4)$$

where $\text{var}(\epsilon) = \sigma^2 I$ and $\text{cov}(\zeta_1, \zeta_2) = \Psi$ and $\text{cov}(\zeta_3 : \zeta_8) = \Phi$.

The parameters and latent variables in the above equations are interpreted as follows.

(i) U_1 is the random intercept of the repeated outcomes.

(ii) U_2 is the random slope of the repeated outcomes.

(iii) λ_j represents the effect of the mediator on the outcome at time j.

(iv) γ_{1j} represents the effect of the main independent variables on the mediator at time j.

(v) κ_j represents the direct effect of the main independent variable on the outcome at time j.

(vi) The product $\lambda_j \times \gamma_{1j}$ represents the indirect effect of the main independent variable on the outcome at time j.

(vii) γ_2 represents the constant effect of the continuous covariate on the repeated outcomes.

The SEM mediation model is represented in Figure 1. In the following diagram we have used the conventions for SEM path diagrams including rectangles representing observed variables, ovals representing latent variables, triangles representing intercept terms, and arrows representing regression relationships between variables.

2.2. SEM Used for Data Generation. The simulated mediated data for this study are generated from an SEM, because our goal is to evaluate the performance of the LMM in a setting where the SEM is assumed to be optimal. We considered a setting where the effects of alcohol, the main independent variable, changed across time. Specifically, we assumed a constant short-term effect of alcohol on CD4 count for the first three time points and a constant long-term effect of alcohol across the last three time points. We refer to this as a "delayed effect" of the main independent variable. To model this delayed effect, we allowed κ_j in (1) to vary. Specifically, we set the first three κ's to be equal ($\kappa^* = \kappa_1 = \kappa_2 = \kappa_3$) and the last three κ's to be equal ($\kappa' = \kappa_4 = \kappa_5 = \kappa_6$), where $\kappa' > \kappa^*$. Short and long-term effects were similarly defined for λ_j and γ_{1j}. Under these assumptions, it can be shown that the predictive formula for a given outcome at time t_j, for $j = 1, 2, 3$ is

$$Y_j = (\alpha_1 + \lambda^* \alpha_3) + \gamma_2 w + \alpha_2 t_j + (\lambda^* \gamma_1^* + \kappa^*) z_j + \left(\zeta_1 + \lambda^* \zeta_{2+j}\right) + \zeta_2 t_j + \epsilon_j, \quad (5)$$

and for $j = 4, 5, 6$ is

$$Y_j = (\alpha_1 + \lambda' \alpha_3) + \gamma_2 w + \alpha_2 t_j + (\lambda' \gamma_1' + \kappa') z_j + \left(\zeta_1 + \lambda' \zeta_{2+j}\right) + \zeta_2 t_j + \epsilon_j. \quad (6)$$

The model assumes a linear effect of time on the outcome.

2.2.1. Simulation Procedures. For the initial set of simulations, we varied the distribution of the total effect of the predictor on the outcome. We evaluated three situations: (i) the total effect was equally distributed between the direct and indirect effect, (ii) the total effect was primarily direct (i.e., the direct effect was larger than the indirect effect through the mediator), and (iii) the total effect was primarily indirect (i.e., the indirect effect through the mediator was larger than the direct effect of the predictor on the outcome).

These simulations considered a setting where the true total effect of the primary independent variable was small (0.05) for the first three time points and small to moderate for the second three time points (0.25), as defined by Cohen [30]. These effect sizes were selected as they are considered feasible and realistic for a wide range of clinical settings. Effect size was defined as the true value of the regression parameter divided by the true standard deviation of the residual error term (ϵ_{ij}). We fixed the true standard deviation of all residual error terms in the simulated data to one, so the effect size is equal to the true value of the regression coefficient. We used a sample size of 350 as this sample size yielded adequate power for the second three time points with the effect size we assumed.

In addition to the initial set of simulations, we also performed simulations evaluating sample sizes ranging from 100–500 and alternative effect sizes, for example, small negative effect sizes as observed in the example data set described in Section 4, a moderate effect size (0.50) as defined by Cohen [30], and a null effect size to evaluate the Type I error properties of the models.

Model performance with respect to the effect of the primary independent variable on the outcome was evaluated separately for each time-point.

We generated data using the SEM described above with repeated measures of a continuous outcome, a random intercept and slope and a time-varying main independent predictor and mediating variable. The outcome, main independent variable, and mediator were each assessed at 6 time points. The following steps were taken to generate the mediated longitudinal data.

(1) Two multivariate normal random variates were generated, one corresponding to the residual variance of the latent intercept and one to be the residual variance of the latent slope.

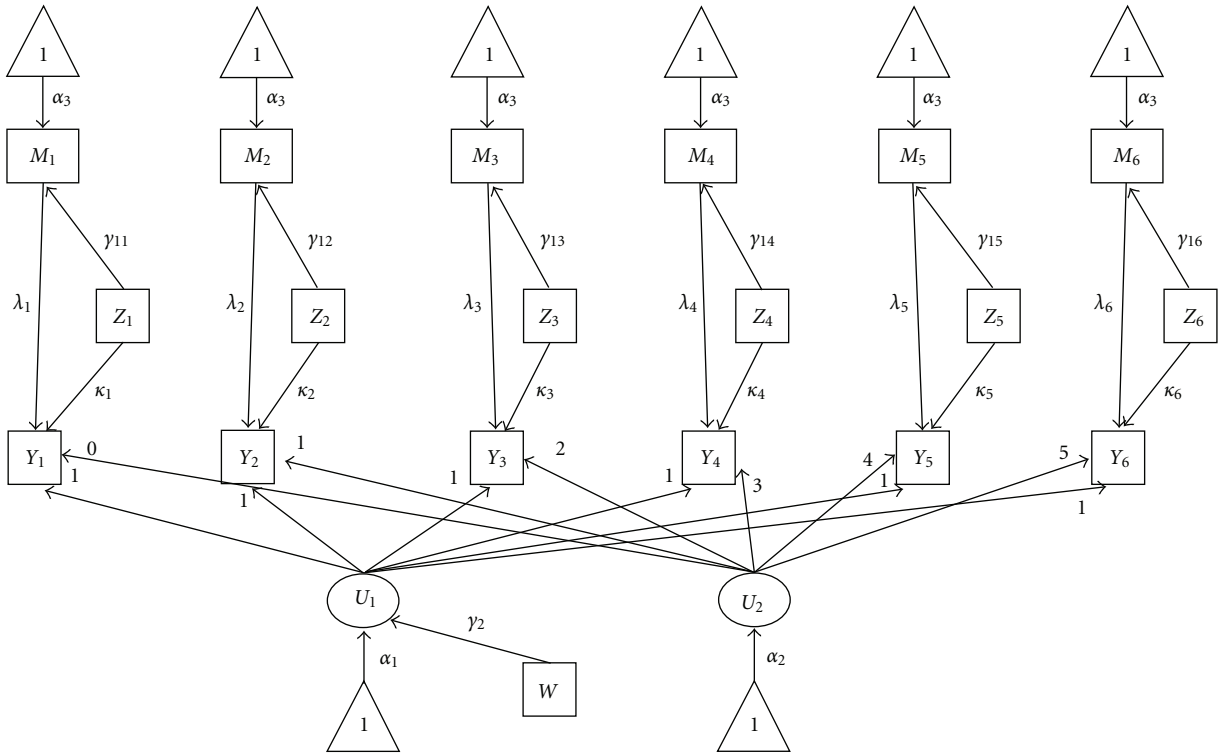


Figure 1: Path diagram of unrestricted structural equation model.

(2) Six multivariate normal random variates were generated corresponding to the residual variance of the mediator variables.

(3) Based on equations (2) and (3), the value of the latent intercept and latent slope were computed.

(4) Based on equation (4), the values of the mediator variables were determined.

(5) Six independent standard normal random variates were generated corresponding to the residual error of the six Y_j's.

(6) Based on equation (1), the value of the Y_j's were generated.

(7) Steps (1) through (6) were repeated 1000 times to create 1000 datasets for each simulation.

The models fit to the simulated data were evaluated by assessing: (i) Bias: estimated as the difference between the true parameter value and the mean observed parameter value divided by the true parameter value. (ii) Coverage probability: estimated as the percentage of the 1000 95% confidence intervals that contained the true parameter value. (iii) Power: estimated as the percentage of the 1000 datasets in which a hypothesis test of the parameter of interest was statistically significant and (iv) Type I error: for settings assuming null effects (for both direct and indirect effects), Type I error was estimated as the percentage of the 1000 datasets in which a hypothesis test of the parameter of interest was statistically significant.

2.3. SEMs and LMMs Fit to the Simulated Data. After the simulated data were generated as described above, the data were fit with three SEMs and five LMMs representing a range of possible models that could be fit to mediated longitudinal data.

2.3.1. Constant Effect SEM. The first SEM we evaluated represents one of the simplest and most common models that can be fit. This model assumes that the direct effect of alcohol on CD4 count is constant (i.e., $\kappa = \kappa_1 = \cdots = \kappa_6$), the effect of alcohol on ART adherence is constant (i.e., $\gamma_{11} = \gamma_{12} = \cdots = \gamma_{16}$), and the effect of ART adherence on CD4 count is constant (i.e., $\lambda_1 = \lambda_2 = \cdots = \lambda_6$). The total effect of the repeated primary independent variable on the repeated outcome is therefore represented by $\kappa + \lambda\gamma_1$. We refer to this model as the constant effect SEM (CESEM).

2.3.2. Delayed Effect SEM. The second SEM fit to the simulated data is the model that was used to simulate the data and defined in Section 2.2; that is, it assumes an early versus late effect. In this model, a short-term total effect of alcohol on CD4 count ($\kappa^* + \lambda^*\gamma_1^*$) is assumed for the first three time points, and a long-term effect of alcohol on CD4 count is assumed for the second three time points ($\kappa' + \lambda'\gamma_1'$).

2.3.3. Unrestricted SEM. The last SEM evaluated is the unrestricted model defined in (1)–(4) and represented in Figure 1. The unrestricted SEM is a model that could be used

to evaluate the nature of a mediated longitudinal relationship between alcohol and HIV disease progression without assuming how the effects may change across time.

2.3.4. Constant Effect Linear Mixed Model. The first mixed model fit to the simulated data assumes the effect of the repeated primary independent variable to remains constant over time. The formula for this constant effect mixed model is

$$Y_j = \beta_0 + \beta_1 w + \beta_2 t_j + \beta_3 z_j + b_1 + b_2 t_j + \epsilon_j, \quad (7)$$

where $\text{var}(\epsilon) = \sigma^2 I$ and $\text{cov}(\mathbf{b}) = \Psi$.

In this model, the interpretation of the parameters is as follows.

(i) β_0 is the intercept of the repeated outcomes.

(ii) β_1 is the effect of the continuous covariate, w, on the repeated outcomes.

(iii) β_2 is the effect of time, t_j, on the repeated outcome.

(iv) β_3 is the effect of the repeated primary independent variable, z_j, on the repeated outcomes.

(v) b_1 is the random intercept of the repeated outcomes.

(vi) b_2 is the random slope of the repeated outcomes.

We note that the mediating variable has been excluded from this model, since the goal is to evaluate the total effect of the main independent variable. If a known mediator is included in a model, then the parameter estimate associated with the primary predictor estimates the direct, rather than the total effect, of that predictor on the outcome [28]. Under the constant effect LMM defined in (7), the total effect of alcohol on CD4 count at any time-point is represented by β_3.

2.3.5. Full Delayed Effect Mixed Model. To capture potential short-term and long-term effects, we allowed the effect of alcohol at the first three time points to differ from that at the last three time points. To accomplish this, an indicator variable representing observations from the last three time points was entered into the model (i.e., indicator variable $I(j > 3) = 1$ at time points 4, 5 and 6 and $I(j > 3) = 0$ otherwise) and the following model was fit:

$$\begin{aligned} Y_j = {} & \beta_0 + \beta_1 w + \beta_2 t_j + \beta_3 I(j > 3) \\ & + \beta_4 z_j + \beta_5 I(j > 3) z_j + b_1 + b_2 t_j + \epsilon_j. \end{aligned} \quad (8)$$

Therefore, the regression model for $j = 1, 2, 3$ would be

$$Y_j = \beta_0 + \beta_1 w + \beta_2 t_j + \beta_4 z_j + b_1 + b_2 t_j + \epsilon_j, \quad (9)$$

and for $j = 4, 5, 6$, it would be

$$Y_j = (\beta_0 + \beta_3) + \beta_1 w + \beta_2 t_j + (\beta_4 + \beta_5) z_j + b_1 + b_2 t_j + \epsilon_j \quad (10)$$

In this model, the total effect of the repeated primary independent variable is represented by β_4 for the first three time points and $\beta_4 + \beta_5$ for the second three time points. In addition, the intercept of the repeated outcomes is given by β_0 for the first three time-periods and by $\beta_0 + \beta_3$ for the second three time-periods. Thus, this model allows for (a) estimating a potentially delayed effect of alcohol (z_j) and (b) accounting for a period effect, by allowing different intercept values for the early and late time periods. The period effect may be induced by the mediator's changing direct effect (in the SEM from which the data are generated, the mediator effect is $\lambda^* \alpha_3$ from (5) for the first three time points and $\lambda' \alpha_3$ from (6) for the last three time points).

2.3.6. Naive Delayed Effect Mixed Model. As described above, the simulated data are generated from an SEM where the effect of the mediator changes over time. In practice, such time dependent effects can be modeled directly as part of the mediation process using SEMs. In contrast, in LMM models, this difference in mean outcome value for early versus late effects can be captured by a time-varying intercept. However, the need for a time-varying intercept term is not immediately clear when fitting a mixed model in this setting, and thus, a model without time-varying intercepts may be more commonly fit. We refer to such a model as the naive delayed effect model

$$\begin{aligned} Y_j = {} & \beta_0 + \beta_1 w + \beta_2 t_j + \beta_3 z_j \\ & + I(j > 3) \beta_4 z_j + b_1 + b_2 t_j + \epsilon_j. \end{aligned} \quad (11)$$

This model is similar to the full delayed model but assumes the intercept of the repeated outcomes, β_0, is the same for all six time periods. In this naive model, the total effect of alcohol on HIV disease progression is given by β_3 for the first three time points and by $\beta_3 + \beta_4$ for the second three time points.

2.3.7. Time Interaction Linear Mixed Model. In mixed models, an interaction between time and the main independent variable is commonly included to assess whether the effect of the independent variable changes linearly across time

$$Y_j = \beta_0 + \beta_1 w + \beta_2 t_j + \beta_3 z_j + \beta_4 t_j z_j + b_1 + b_2 t_j + \epsilon_j. \quad (12)$$

In this model, the total effect of alcohol (z_j) is modeled as a linear function of time, t_j, and is represented by $\beta_3 + \beta_4 t_j$.

2.3.8. Unrestricted Mixed Model. The last mixed model we evaluated allowed the effect of alcohol on CD4 count to vary at each time-point, without assuming linearity. The equation for this unrestricted LMM is

$$\begin{aligned} Y_j = {} & \beta_0 + \beta_w w + \beta_{t_j} I\left(t_j\right) + \beta_z z_j + \beta_{z,t_j} I\left(t_j\right) z_j \\ & + b_1 + b_2 t_j + \epsilon_{ij} \end{aligned} \quad (13)$$

where $I(t_j)$ is an indicator of time point and is defined as $I(t_j) = 1$ if $t_j = j$ and $I(t_j) = 0$ otherwise. In this model, the effect of z_j is a function of time and is represented by $\beta_z + \beta_{z,t_j} I(t_j)$. This is the least restrictive model and is sometimes called a profile analysis [31].

2.4. Model Comparisons. To evaluate the performance of the LMM relative to the SEM, we made the following five comparisons.

(1) Constant effect SEM (CESEM) versus constant effect mixed model (CEMM).

(2) Delayed effect SEM (DESEM) versus full delayed effect mixed model (FDEMM).

(3) Unrestricted SEM (USEM) versus unrestricted mixed model (UMM).

(4) Delayed effect SEM (DESEM) versus naive delayed effect mixed model (NDEMM).

(5) Unrestricted SEM (USEM) versus time interaction mixed model (TIMM).

We simulated data under the SEM defined in Section 2.2. The SEMs were fit as a reference standard to compare with the LMM results, since the objective was to evaluate the performance of the LMM in a setting where the SEM is assumed to be optimal. For comparisons (1), (2), and (3), the main difference between the models is that the SEM explicitly models the mediation, while the LMM does not. All other aspects of the model are the same. Comparison 4 is of interest, because with time-varying mediated data, the naive delayed effect model is commonly fit within the mixed model framework. However, as described earlier, this model does not fully capture the time-varying mediation process, and thus, it is useful to evaluate its performance against the SEM. Comparison (5) is evaluated since a time interaction mixed model is also a common approach in the mixed model framework when a time-varying relationship is suspected. However, it relies on the assumption that the effect of alcohol is a linear function of time. It is, therefore, of interest to compare this model to the unrestricted SEM, which does not assume linearity.

3. Results of Simulation Study

3.1. Constant Effect SEM versus Constant Effect Mixed Model. In a setting where the true effect size changed over time, the estimated power to detect the true effect of the primary independent variables on the outcome from a model assuming a constant effect was generally inadequate with a sample size of 350 ($\leq$66% for both the SEM and LMM in all cases) (Table 1). When effects were distributed equally between direct and indirect effects, estimated power was similar for the two models although slightly higher for the SEM (65% versus 62%). The bias estimates for both the CESEM and CEMM were quite large (180% and 171%, resp., for t_1–t_3 and -44% and -45%, resp., for t_4–t_6), overestimating smaller short-term effects and underestimating larger long-term effects as would be expected. The coverage probability was also quite low although for both models, it was higher for the early time points compared to the later three time points. Similar results were observed when effects were primarily direct and also when they were primarily indirect. We note that we deliberately created a small effect at the first three time points to simulate a delayed effect of treatment on outcome and, therefore, did not expect to have adequate power to detect effects at the first three time points with the sample size evaluated. Similar patterns were observed with different sample sizes and effect sizes. Power was markedly lower for sample sizes less than 350 and for the reduced effect sizes.

3.2. Delayed Effect SEM versus Full Delayed Effect Mixed Model. The DESEM and FDEMM had similar power to detect long-term total effects independent of whether effects were equally distributed, primarily direct, and primarily indirect (Table 2). With a sample size of 350 and an effect size of 0.25, the estimated power for the last three time points for the DESEM was slightly higher (83%–85%) than for the FDEMM (82%–84%). The bias for both models was low (-0.1%-1.7% and -0.3%-1.4%, resp.) and the coverage probability was high (95% and 94% for the DESEM and FDEMM, resp.). Similar patterns were observed for other sample sizes with the same effect size. For smaller sample sizes (100 and 200), the power dropped to unacceptable levels (32%–63%).

Again, since the magnitude of the effect at the first three time points is small, we did not expect to have adequate power to detect such an effect in either modeling framework with a sample size of 350. In both models, the power to detect the total effect for the first three time periods was substantially lower than that for the last three (10%–13% in the first three time points versus 82%–85% for the second three time points for both models), where the effects were of a larger magnitude. With a sample size of 400, the power remained low to detect a small effect (-0.11, the observed effect size from the real data example standardized by the residual standard deviation) for both models (30% for DESEM and FDEMM). At all sample and effect sizes, results did not differ substantially between modeling frameworks.

3.3. Unrestricted SEM versus Unrestricted Mixed Model. With a sample size of 350, the performance of the USEM and UMM were very similar, regardless of whether effects were equally distributed, primarily direct, or primarily indirect (Table 3). As seen in previous models, the power to detect the effect at the first three time points was low (6%–9%) for both models. For the last three time points, the power to detect the effects was also low for both models (36%–58% for the USEM and 36%–55% for the UMM). The bias, however, was also quite low for both models (-0.08% to 2.9% for the USEM and -2.6% to 2.6% for the UMM). The coverage probability for both models was good (93%–96% for the USEM and UMM) across the different effect distributions. In these models, no specific relationship with time is assumed in the LMM or the SEM, so both models freely estimate the effect of the time-varying main independent variable on time. The cost of this, however, is that several more parameters must be estimated, and therefore, the power to detect effects is reduced. Similar patterns were observed for other sample sizes and effect sizes.

3.4. Delayed Effect SEM versus Naive Delayed Effect Mixed Model. With the sample size of 350, when the distribution of the effect was equally distributed, the power to detect the

Table 1: Performance of SEM and linear mixed model assuming total effect of main independent variable is constant when true underlying effects are small for early time points and small to moderate for late time points.

Simulated data scenarios				Constant effect SEM			Constant effect LMM		
Time point	Effect size	Sample size	Effect distribution	Bias (%)	Coverage probability (%)	Power (%)	Bias (%)	Coverage probability (%)	Power (%)
$t_1 : t_3$	0.05	350	Equal	180	67	65	171	71	62
$t_4 : t_6$	0.25	350	Equal	−44	56	65	−45	53	62
$t_1 : t_3$	0.05	350	Primarily indirect	179	70	60	164	73	55
$t_4 : t_6$	0.25	350	Primarily indirect	−44	55	60	−47	51	55
$t_1 : t_3$	0.05	350	Primarily direct	173	69	66	169	69	65
$t_4 : t_6$	0.25	350	Primarily direct	−45	50	66	−46	49	65
$t_1 : t_3$	0.05	100	Equal	168	86	24	159	88	22
$t_4 : t_6$	0.25	100	Equal	−46	82	24	−48	81	22
$t_1 : t_3$	0.05	200	Equal	178	80	41	169	81	38
$t_4 : t_6$	0.25	200	Equal	−44	73	41	−46	70	38
$t_1 : t_3$	0.05	400	Equal	175	65	67	165	69	64
$t_4 : t_6$	0.25	400	Equal	−45	49	67	−47	46	64
$t_1 : t_3$	0.05	350	Equal	395	3	97	369	2	94
$t_4 : t_6$	0.5	350	Equal	−50	16	97	−53	21	94
$t_1 : t_3$	0.05	400	Equal	−148	73	8	−137	76	6
$t_4 : t_6$	−0.11	400	Equal	−78	63	8	−83	60	6

Based on 1000 simulated datasets.

total effect for last three time points was very good for the DESEM (83%), the bias was low (−0.1%), and the coverage probability was high (95%) (Table 2). In contrast, for the NDEMM, there was substantial bias (109%) in estimating the total effect for the last three time points. This naive model clearly does not correctly estimate the effect of the primary independent variable on the outcome. Similar trends were observed in the comparison of the two models regardless of how the total effect was distributed, the sample size, or the effect size.

3.5. Unrestricted SEM versus Time-Interaction Mixed Model. Regardless of the distribution of effects, sample size, or effect size, the TIMM and USEM had low power to detect the effect of the repeated primary independent variable on the repeated outcome for the first three time points (Table 3), as expected. With a sample size of 350, power ranged 7%–9% for the USEM and 5%–56% for the TIMM. For the last three time points at this sample size, the USEM had lower bias but also lower power compared to the TIMM. For the TIMM, incorrectly forcing a linear trend resulted in a larger degree of bias. While the TIMM had a large degree of bias at all time points, it has higher power than the USEM at most time points. This increased power relative to the USEM is likely due, at least in part, to fewer parameters being estimated. The higher power and bias of the TIMM relative to the USEM was also observed for other sample sizes and effect sizes (Table 3).

3.6. Type I Error Rates. Table 4 shows the estimated Type I error rates for a range of sample sizes. The nominal Type I error rate was 0.05. The Type I error was remarkably similar between analogous SEM and LMM models. Across all models, the observed Type I error rates ranged from 0.030 (CESEM, sample size of 350) to 0.072 (UMM, sample size of 100).

4. Real-Data Example: Alcohol and HIV Disease Progression

To demonstrate the application of the various LMMs and SEMs evaluated in the simulation study, we analyzed data from a prospective cohort study evaluating the effect of alcohol use on HIV disease progression. Samet et al. have previously reported the analyses from this longitudinal cohort study [29]. The original analyses combined data from two cohorts (the HIV-ALC and HIV-LIVE cohorts). To illustrate the models evaluated in this paper, we have used data from the HIV-LIVE study and fit the various LMMs and SEMs of interest. For clarity of presentation, we limited the analyses to observations where subjects reported any ART use during followup ($n = 319$) and included only the following key variables: heavy alcohol consumption (yes versus no), the main independent variable, ART adherence (percentage of pills taken in the last three days), the mediator;

Table 2: Performance of SEM and linear mixed model assuming delayed effects of main independent variable when true underlying effects are small for early time points and small to moderate for late time points.

Simulated data scenarios			Delayed effect SEM			Naive delayed effect LMM			Full delayed effect LMM		
Time point	Effect size	Sample size	Bias (%)	Coverage probability (%)	Power (%)	Bias (%)	Coverage probability (%)	Power (%)	Bias (%)	Coverage probability (%)	Power (%)
$t_1 : t_3$	0.05	350	10	95	13	390	23	53	9.6	95	13
$t_4 : t_6$	0.25	350	−0.1	95	83	109	9	100	−0.3	94	82
$t_1 : t_3$	0.05*	350	−4.2	95	11	522	6	82	−5.0	95	10
$t_4 : t_6$	0.25*	350	1.7	95	83	144	0.8	100	1.4	94	82
$t_1 : t_3$	0.05**	350	2.2	94	11	−234	61	17	2.2	94	12
$t_4 : t_6$	0.25**	350	−0.7	95	85	633	48	100	−0.7	96	84
$t_1 : t_3$	0.05	100	−0.2	94	8	−397	68	20	−0.9	94	7
$t_4 : t_6$	0.25	100	−2.9	94	34	107	58	92	−2.5	94	32
$t_1 : t_3$	0.05	200	0.2	95	10	−396	45	34	-0.9	95	9
$t_4 : t_6$	0.25	200	0.8	95	63	109	28	99	0.7	96	61
$t_1 : t_3$	0.05	400	0.2	94	14	−400	17	60	−1.4	94	12
$t_4 : t_6$	0.25	400	8.0	96	88	109	6	100	−0.8	94	86
$t_1 : t_3$	0.05	350	4.2	96	11	−684	0.2	97	4.1	96	9
$t_4 : t_6$	0.5	350	0.5	95	100	98	0	100	0.09	95	100
$t_1 : t_3$	0.05	400	0.4	95	11	−148	80	6	0.6	95	10
$t_4 : t_6$	−0.11	400	1.5	95	30	−90	74	6	2	95	30

Based on 1000 simulated datasets.
Results are from simulated data with total effects equally distributed between direct and indirect effects, except where indicated.
*Total effect is primarily direct.
**Total effect is primarily indirect.

age, a potential confounder, and CD4 cell count, the primary outcome. Each variable was assessed every six months for up to four years.

The total effect of alcohol consumption on CD4 count was not statistically significant in any of the SEMs or LMMs fit to the data. Estimated total effects are detailed in Table 5. Both constant effect models showed a small negative effect (−3.7 in the CESEM and −3.0 in the CEMM). The delayed effect SEM and LMM showed similar negative effects in the last four time points although the magnitude of effect in the DESEM was slightly larger (−10.3) than that for the DEMM (−4.1). The magnitude of effect at the first three time points was quite small in both delayed effect models but differed in sign in the DESEM (0.41) and DEMM (−2.3) although neither value was significantly different from zero. The unrestricted models generally showed similar results with effects ranging from −41.8 to 5.7 in the USEM and ranging from −15.8 to 9.5 in the UMM. The direction of the estimated alcohol effects were consistent between models with the exception of the third time-point which had a small estimated negative effect in the USEM (−1.4) and a small estimated positive effect in the UMM (6.9); however, neither effect was statistically significant. The magnitude of the effects were similar between the TIMM and USEM. Since a linear effect of time is assumed in the TIMM, however, all effects after time-point 2 are negative, whereas in the USEM, the direction of effects changes between negative and positive.

5. Discussion

Mixed models are a useful technique to analyze longitudinal data, with time-dependent variables. They can be applied to mediated longitudinal data, and a series of models can be fit to disentangle direct versus indirect effects of an exposure. However, it is unknown whether they perform well relative to SEMs, a method used for mediational analysis. In this paper, we evaluated the performance of the linear mixed model relative to the SEM in the setting of a time-dependent predictor and mediator, where the effects of both change over time.

The main simulation study assumed that the primary independent predictor had a delayed effect on the outcome (i.e., a small effect at the first three time points and a moderate effect at the last three time points). A range of SEMs (constant effect SEM, delayed effect SEM, and unrestricted SEM) and LMMs (constant effect mixed model, naive delayed effect mixed model, full delayed effect mixed model, time-interaction mixed model, and unrestricted mixed model) were fit to the simulated data.

Three comparisons were made between "analogous" models in that the main difference between models was that the SEM explicitly models the mediation, while in the mixed model, the mediator is removed from the model. The analogous models were constant effect SEM versus constant effect mixed model delayed effect SEM versus full delayed effect mixed model, and unrestricted SEM versus unrestricted mixed model. For each of the three comparisons,

Table 3: Performance of unrestricted structural equation model (USEM) and unrestricted and time interaction linear mixed models (UMM and TIMM, resp.) when true underlying effects are small for early time points and small to moderate for late time points.

Simulated data scenarios			USEM			TIMM			UMM		
Time point	Effect size	Sample size	Bias (%)	Coverage probability (%)	Power (%)	Bias (%)	Coverage probability (%)	Power (%)	Bias (%)	Coverage probability (%)	Power (%)
t_1	0.05	350	16	95	8	−50	93	6	16	96	7
t_2	0.05	350	9.4	94	8	51	94	19	8.6	94	8
t_3	0.05	350	7.4	94	9	153	75	56	4.9	95	8
t_4	0.25	350	−0.6	94	54	−29	81	76	−0.09	94	54
t_5	0.25	350	−0.4	94	48	−8.9	94	73	−0.5	93	48
t_6	0.25	350	−0.08	93	40	11	93	66	−1.0	94	38
t_1	0.05*	350	−4.0	94	8	−71	94	5	−6.3	95	7
t_2	0.05*	350	−9.6	96	7	38	94	15	−9.8	95	6
t_3	0.05*	350	−1.4	95	8	146	78	49	1.3	95	7
t_4	0.25*	350	2.9	95	55	−29	81	70	2.6	94	54
t_5	0.25*	350	2.2	95	49	−7.4	94	71	1.6	94	48
t_6	0.25*	350	0.4	95	36	14	94	66	−0.5	95	36
t_1	0.05**	350	10	95	7	−60	94	6	8.7	95	7
t_2	0.05**	350	−14	93	7	44	95	18	−14	93	7
t_3	0.05**	350	11	94	8	148	75	56	12	94	9
t_4	0.25**	350	−0.4	95	58	−29	81	79	−0.3	95	55
t_5	0.25**	350	0.0	96	50	−8.6	95	77	0.2	96	48
t_6	0.25**	350	−0.2	96	37	12	95	70	−2.6	96	36
t_1	0.05	100	−25	94	7	−26	94	6	−69	93	6
t_2	0.05	100	8.8	94	6	7.2	93	6	36	94	9
t_3	0.05	100	10	96	5	14	95	5	141	90	20
t_4	0.25	100	−2.0	94	20	−2.2	93	20	−31	90	30
t_5	0.25	100	−3.8	94	17	−3.2	94	17	−9.8	95	29
t_6	0.25	100	−2.5	94	16	−2.2	94	16	11	95	26
t_1	0.05	200	8.2	95	7	−60	95	5	7.7	95	7
t_2	0.05	200	1.2	95	7	45	95	13	0.4	95	6
t_3	0.05	200	−7.8	95	7	150	83	34	−9.2	95	6
t_4	0.25	200	4.0	96	38	−29	97	52	4.1	95	36
t_5	0.25	200	−3.0	95	28	−8.1	95	50	−2.6	95	28
t_6	0.25	200	−0.7	94	24	13	95	45	−1.3	94	25
t_1	0.05	400	41	95	8	−65	94	5	2.7	95	7
t_2	0.05	400	−3.2	94	7	41	94	19	−4.7	97	6
t_3	0.05	400	−0.2	95	9	146	74	58	−2.3	96	7
t_4	0.25	400	−1.3	95	60	−30	76	80	−1.7	95	58
t_5	0.25	400	1.2	95	54	−8.6	95	79	1.0	94	53
t_6	0.25	400	−0.5	96	43	12	95	73	−1.9	96	40
t_1	0.05	400	−6	95	7	52	94	14	−7	95	7
t_2	0.05	400	4	97	6	−33	95	8	3	97	7
t_3	0.05	400	2	95	8	−119	82	5	5	95	7
t_4	−0.11	400	0.4	95	17	−52	83	14	0.9	95	17
t_5	−0.11	400	2.5	95	17	−13	95	23	4	95	17
t_6	−0.11	400	1	95	13	26	94	26	2	95	13

Based on 1000 simulated datasets.
Results are from simulated data with total effects equally distributed between direct and indirect effects, except where indicated.
*Total effect is primarily direct.
**Total effect is primarily indirect.

Table 4: Type I error rates for mediated structural equation models and linear mixed models at various sample sizes.

Simulated data		Unrestricted		Delayed effect		Constant effect		Time interaction
Time point	Sample size	SEM	LMM	SEM	LMM	SEM	LMM	LMM
t_1	100	0.050	0.050	0.053	0.053	0.05	0.053	0.053
t_2	100	0.064	0.072	0.053	0.053	0.05	0.053	0.065
t_3	100	0.038	0.039	0.053	0.053	0.05	0.053	0.053
t_4	100	0.047	0.051	0.045	0.043	0.05	0.053	0.045
t_5	100	0.048	0.051	0.045	0.043	0.05	0.053	0.041
t_6	100	0.050	0.053	0.045	0.043	0.05	0.053	0.038
t_1	350	0.053	0.053	0.048	0.046	0.030	0.031	0.054
t_2	350	0.044	0.044	0.048	0.046	0.030	0.031	0.047
t_3	350	0.051	0.051	0.048	0.046	0.030	0.031	0.032
t_4	350	0.049	0.049	0.046	0.045	0.030	0.031	0.038
t_5	350	0.038	0.038	0.046	0.045	0.030	0.031	0.046
t_6	350	0.052	0.052	0.046	0.045	0.030	0.031	0.049
t_1	500	0.058	0.064	0.047	0.045	0.054	0.053	0.052
t_2	500	0.048	0.048	0.047	0.045	0.054	0.053	0.049
t_3	500	0.049	0.046	0.047	0.045	0.054	0.053	0.054
t_4	500	0.049	0.050	0.048	0.048	0.054	0.053	0.047
t_5	500	0.051	0.048	0.048	0.048	0.054	0.053	0.048
t_6	500	0.046	0.047	0.048	0.048	0.054	0.053	0.044

Based on 1000 simulated datasets.

Table 5: The total effect of heavy alcohol consumption on CD4 cell count from a prospective cohort study of HIV-infected subjects on antiretroviral therapy ($n = 319$) [29]. Longitudinal regression analyses were performed using linear mixed models and structural equation models, and adjusted mean differences (SE) are reported.

	SEM			LMM			
Time point	Constant effect	Delayed effect	Unrestricted	Constant effect	Delayed effect	Time interaction	Unrestricted
t_1	−3.7 (9.6)	0.41 (11.6)	−4.8 (19.6)	−3.0 (11.3)	−2.3 (13.8)	0.44 (22.0)	−7.7 (23.9)
t_2	−3.7 (9.6)	0.41 (11.6)	0.13 (20.1)	−3.0 (11.3)	−2.3 (13.8)	−0.41 (18.1)	6.2 (24.1)
t_3	−3.7 (9.6)	0.41 (11.6)	−1.4 (20.4)	−3.0 (11.3)	−2.3 (13.8)	−1.3 (14.7)	6.9 (23.7)
t_4	−3.7 (9.6)	0.41 (11.6)	5.7 (18.9)	−3.0 (11.3)	−2.3 (13.8)	−2.1 (12.3)	0.68 (22.3)
t_5	−3.7 (9.6)	−10.3 (14.0)	−3.4 (20.2)	−3.0 (11.3)	−4.1 (16.5)	−3.0 (11.3)	−4.6 (24.1)
t_6	−3.7 (9.6)	−10.3 (14.0)	3.8 (22.0)	−3.0 (11.3)	−4.1 (16.5)	−3.8 (12.3)	9.5 (25.1)
t_7	−3.7 (9.6)	−10.3 (14.0)	−13.1 (25.9)	−3.0 (11.3)	−4.1 (16.5)	−4.7 (14.8)	−13.6 (30.3)
t_8	−3.7 (9.6)	−10.3 (14.0)	−41.8 (31.6)	−3.0 (11.3)	−4.1 (16.5)	−5.5 (18.2)	−15.8 (35.1)

the SEM and LMM yielded similar results. The power, bias, and coverage probability were all similar when the SEM and LMM were compared. The results from the analysis of data from a prospective cohort study evaluating the impact of alcohol use on HIV disease progression further illustrated the similarity of results from analogous SEMs and LMMs.

We also considered two comparisons of nonanalogous models. The first comparison was between the delayed effect SEM and the naive delayed effect mixed model. In the SEM framework, mediation can be directly modeled at each time-point, and therefore, the mediated delayed effect of the time-varying predictor is easily incorporated. In the mixed model framework, however, mediation is not directly modeled. Instead, mediators are removed from the model if the goal is to obtain the total effect of the time-varying predictor on the outcome [28]. Therefore, in the mixed model framework, it may not be clear whether a time-varying intercept term is necessary in the model to properly account for the mediated relationship between the predictor and outcome. Our simulations show that the naive delayed mixed model produced extremely biased estimates of both short- and long-term exposure effects, and coverage probabilities were poor. Therefore, although the naive delayed effect mixed model represents a model that may be a natural choice in the mixed model framework, it may not produce valid estimates. To obtain accurate estimates with the mixed model, fitting the full delayed effect model (with time-specific intercept terms) was required. However, as noted earlier, this model may be nonintuitive. This is a distinct disadvantage of the mixed model framework since the model that may be the most natural to fit may result in inaccurate estimates, whereas a natural choice for the SEM is the full delayed effect model, a model which performed relatively well. The second set of nonanalogous models compared the unrestricted SEM

and the time-interaction mixed model. These two models reflect a potential difference in the way that time is handled in the two frameworks. In longitudinal data analysis, SEMs incorporate the value of time as a fixed regression coefficient in the measurement model. Treatment of time is usually limited to a linear main effect of time. If some unspecified nonlinear relationship over time between the predictor and outcome is suspected, the most natural way to evaluate this is to leave the relationship between the time-varying predictor and outcome unrestricted and obtain separate estimates at each time-point as is done in the unrestricted SEM. In mixed models, however, interactions between time and other predictors (time invariant or time varying) are frequently incorporated. In our simulation study, the time interaction mixed model had substantially larger bias compared to the unrestricted SEM. Power was generally higher for the mixed model, possibly due in part to the fewer number of parameters being estimated. The difference between the time-interaction mixed model and the unrestricted SEM was also observed in the real-data example.

In the setting of mediated longitudinal data where exposure effects change over time, the mixed model performed well relative to analogous SEMs. The delayed effect SEM and full delayed effect mixed model had the best performance in terms of bias, coverage probability, and power in modeling the time-specific relationships between variables. It should be noted that in the setting of mediated time-specific effects, the delayed effect SEM, a natural choice for a model within the SEM framework, yielded substantially better results than the naive delayed mixed model, a natural model to choose within the LMM framework. Two other common models that may be fit, the unrestricted SEM and mixed model, both performed well in terms of bias and coverage probability, however, both had lower power due to the relatively large number of parameters being estimated for the given sample size. We note that the results observed in this study may not be generalizable to other settings, for example, scenarios with more complex pathways and relationships between variables could affect the performance of the LMM.

Linear mixed models can perform well relative to SEMs in the analysis of mediated longitudinal data with a time-dependent predictor and mediator. However, care must be taken to identify an appropriate model that adequately accounts for mediator effects, for example, by including time-varying intercepts and excluding variables in the causal pathway. In the specific setting of delayed effect of the time-varying predictor, common models fit within the mixed model framework may not perform adequately in this mediated longitudinal data setting. However, an appropriately specified mixed model can have good performance relative to the SEM in evaluating the overall effects of a time-varying predictor.

Acknowledgments

The authors wish to thank Howard Cabral and Timothy Heeren for thoughtful discussions of the problem, and the anonymous journal reviewer for constructive and helpful comments. We also thank Jeffrey H. Samet, PI of the HIV-LIVE Study (NIAAA: R01-AA13216), for use of the data.

References

[1] N. M. Laird and J. H. Ware, "Random-effects models for longitudinal data," *Biometrics*, vol. 38, no. 4, pp. 963–974, 1982.

[2] R. M. Baron and D. A. Kenny, "The moderator-mediator variable distinction in social psychological research. Conceptual, strategic, and statistical considerations," *Journal of Personality and Social Psychology*, vol. 51, no. 6, pp. 1173–1182, 1986.

[3] J. L. Krull and D. P. MacKinnon, "Multilevel mediation modeling in group-based intervention studies," *Evaluation Review*, vol. 23, no. 4, pp. 418–444, 1999.

[4] K. A. Bollen, *Structural Equations with Latent Variables*, John Wiley & Sons, New York, NY, USA, 1989.

[5] B. Muthen, "Analysis of longitudinal data using latent variable models with varying parameters," in *Best Methods for the Analysis of Change: Recent Advances, Unanswered Questions, Future Directions*, pp. 1–17, APA, Washington, DC, USA, 1991.

[6] P. D. Mehta and S. G. West, "Putting the individual back into individual growth curves," *Psychological Methods*, vol. 5, no. 1, pp. 23–43, 2000.

[7] M. J. Rovine and P. C. M. Molenaar, "A structural modeling approach to a multilevel random coefficients model," *Multivariate Behavioral Research*, vol. 35, no. 1, pp. 51–88, 2000.

[8] P. J. Curran, "Have multilevel models been structural equation models all along?" *Multivariate Behavioral Research*, vol. 38, no. 4, pp. 529–569, 2003.

[9] D. J. Bauer, "Estimating multilevel linear models as structural equation models," *Journal of Educational and Behavioral Statistics*, vol. 28, no. 2, pp. 135–167, 2003.

[10] R. D. Stoel, G. van Den Wittenboer, and J. Hox, "Estimating multilevel linear models as structural equation models," *Metodologia de las Ciencias del Comportamiento*, vol. 5, pp. 21–23, 2003.

[11] P. D. Mehta and M. C. Neale, "People are variables too: multilevel structural equations modeling," *Psychological Methods*, vol. 10, no. 3, pp. 259–284, 2005.

[12] B. N. Sánchez, E. Budtz-Jørgensen, L. M. Ryan, and H. Hu, "Structural equation models: a review with applications to environmental epidemiology," *Journal of the American Statistical Association*, vol. 100, no. 472, pp. 1443–1455, 2005.

[13] J. J. McArdle and D. Epstein, "Latent growth curves within developmental structural equation models," *Child Development*, vol. 58, no. 1, pp. 110–133, 1987.

[14] B. O. Muthén and P. J. Curran, "General longitudinal modeling of individual differences in experimental designs: a latent variable framework for analysis and power estimation," *Psychological Methods*, vol. 2, no. 4, pp. 371–402, 1997.

[15] M. J. Rovine and P. C. M. Molenaar, "A structural equations modeling approach to the general linear mixed model," in *New Methods for the Analysis of Change*, pp. 67–96, APA, Washington, DC, USA, 2001.

[16] J. J. McArdle and M. S. Aber, "Patterns of change within latent variable structural equations models," in *Statistical Methods in Longitudinal Research*, pp. 151–224, Academic Press, New York, NY, USA, 1990.

[17] W. Meredith and J. Tisak, "Latent curve analysis," *Psychometrika*, vol. 55, no. 1, pp. 107–122, 1990.

[18] J. B. Willett and A. G. Sayer, "Using covariance structure analysis to detect correlates and predictors of individual change over time," *Psychological Bulletin*, vol. 116, no. 2, pp. 363–381, 1994.

[19] B. Muthen, "Latent variable modeling of longitudinal and multilevel data," in *Sociological Methodology*, pp. 453–480, Blackwell, Washington, DC, USA, 1997.

[20] K. A. Bollen and P. J. Curran, *Latent Curve Models: A Structural Equation Perspective*, John Wiley & Sons, Hoboken, NJ, USA, 2006.

[21] J. Cheong, D. P. MacKinnon, and S. T. Khoo, "Investigation of mediational processes using parallel process latent growth curve modeling," *Structural Equation Modeling*, vol. 10, no. 2, pp. 238–262, 2003.

[22] D. P. MacKinnon, *Introduction to Statistical Mediation Analysis*, Lawrence Erlbaum Associates, New York, NY, USA, 2008.

[23] J. L. Krull and D. P. MacKinnon, "Multilevel modeling of individual and group level mediated effects," *Multivariate Behavioral Research*, vol. 36, no. 2, pp. 249–277, 2001.

[24] S. E. Maxwell and D. A. Cole, "Bias in cross-sectional analyses of longitudinal mediation," *Psychological Methods*, vol. 12, no. 1, pp. 23–44, 2007.

[25] D. A. Kenny, J. D. Korchmaros, and N. Bolger, "Lower level mediation in multilevel models," *Psychological Methods*, vol. 8, no. 2, pp. 115–128, 2003.

[26] D. J. Bauer, K. J. Preacher, and K. M. Gil, "Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations," *Psychological Methods*, vol. 11, no. 2, pp. 142–163, 2006.

[27] D. A. Cole and S. E. Maxwell, "Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling," *Journal of Abnormal Psychology*, vol. 112, no. 4, pp. 558–577, 2003.

[28] E. A. Blood, H. Cabral, T. Heeren, and D. M. Cheng, "Performance of mixed effects models in the analysis of mediated longitudinal data," *BMC Medical Research Methodology*, vol. 10, article 16, 2010.

[29] J. H. Samet, D. M. Cheng, H. Libman, D. P. Nunes, J. K. Alperen, and R. Saitz, "Alcohol consumption and HIV disease progression," *Journal of Acquired Immune Deficiency Syndromes*, vol. 46, no. 2, pp. 194–199, 2007.

[30] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1988.

[31] G. M. Fitzmaurice, N. M. Laird, and J. H. Ware, *Applied Longitudinal Analysis*, John Wiley & Sons, Hoboken, NJ, USA, 2004.

J Immigrant Minority Health (2011) 13:323–332
DOI 10.1007/s10903-008-9217-4

ORIGINAL PAPER

Clinical Issues in Caring for Former Chattel Slaves

Rebecca D. Blumhofer · Neha Shah · Michael A. Grodin · Sondra S. Crosby

Published online: 17 December 2008

Abstract Over the centuries, slavery has become embedded into the social fabric of Mauritania with generations of abid and bizan (Mauritanian slaves and slave masters, respectively) born and raised knowing nothing but the institution of chattel slavery. Abid fleeing their station in Mauritania come to the USA with unique psychological needs that will affect all of their interactions with the medical community. This paper aims to assist health professionals and others concerned with the welfare of former chattel slaves in competently serving this vulnerable population. Discussion includes an overview of Mauritanian chattel slavery, deduced sequelae of chattel slavery, preliminary recommendations for mental health and medical treatment protocols, and suggestions for future research. A confidential Institutional Review Board (IRB)-approved case report will be used to illustrate these objectives.

R. D. Blumhofer
Boston University School of Medicine, Boston, MA 02118, USA

R. D. Blumhofer (✉)
28 Mercer Street #3, South Boston, MA 02127, USA
e-mail: rds81@bu.edu

R. D. Blumhofer · M. A. Grodin
Department of Health Law, Bioethics and Human Rights, Boston University School of Public Health, Boston, MA, USA

R. D. Blumhofer · N. Shah · M. A. Grodin · S. S. Crosby
Boston Medical Center, Boston Center for Refugee Health and Human Rights, Boston, MA, USA

M. A. Grodin
Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA

S. S. Crosby
Department of Medicine, Boston University School of Medicine, Boston, MA, USA

Keywords Chattel slavery · Mauritania · Slave mentality · Vulnerable patient population

Introduction

The Office of the UN High Commissioner for Human Rights asserts that the term "slavery" includes a variety of practices that violate human rights [1]. Some form of slavery is practiced on nearly every continent, involving approximately 27 million enslaved persons [2]. Table 1 describes different kinds of slavery found worldwide. In chattel slavery, slaves are born into a lifetime of bondage [3].

Mauritania is one of the few places on earth where chattel slavery is still practiced. Over the centuries, slavery has become embedded into the social fabric of Mauritania with generations of *abid* and *bizan* (Mauritanian slaves and slave masters, respectively) born and raised knowing nothing but the institution of chattel slavery [4]. An August 2007 Boston Globe article alerted its readers to the persistence of chattel slavery in Mauritania when reporting on the recent passing of a law promising jail time and fines for slaveholders, as well as reparations for those who have been enslaved. "The new law also makes any 'cultural or artistic work defending slavery' punishable by two years in prison, and makes it an offense for governmental authorities not to pursue slaveholders" [5]. Clinicians at the Boston Center for Refugee Health and Human Rights have been aware of Mauritanian chattel slavery because patients who bear the mental and physical scars of Mauritanian chattel slavery seek care and advocacy at the Center, including documentation of their experiences for asylum hearings. These patients have motivated the authors to begin the dialogue on how to sensitively meet the psychological and physical needs of female survivors of chattel slavery. Because Mauritanian chattel slavery is well

Table 1 Forms of slavery occurring worldwide [3, 8]

Forced and bonded labor. According to the International Labor Organization, an estimated 246 million children are engaged in exploitative child labor, with almost three quarters of them working in hazardous environments, such as mines or factories, or with dangerous substances like chemicals and agricultural pesticides. There are several types of forced labor slavery including: chattel, debt and forced labor. Countries involved in this activity include but are not limited to: South Asia, Central American and the United States.
Human trafficking. The use of children as a commodity for labor or sex is a lucrative international trade. An estimated 1.2 million children worldwide are trafficked each year, and some are arrested and detained as illegal aliens. Girls as young as 13, mainly from Asia and Eastern Europe, are trafficked as "mail-order brides". Up to 10,000 women and girls from poor neighboring countries have been lured into commercial sex establishments. Like other forms of criminal activity, trafficking is an underground activity and difficult to address.
Sexual exploitation. About 1 million children, mostly girls but also a significant number of boys, are exploited every year in the multi-billion-dollar sex industry. Such commercial abuse of children is fuelled by local demand, with sexual tourism only a small part of the problem. Since sexual activity is usually regarded as a private matter, Governments and communities alike are often reticent to intervene in cases of sexual exploitation. Countries involved in this activity include but are not limited to: South Asia and the United States.
Child soldiers. At any given time, over 300,000 child soldiers, as young as 8, are exploited in armed conflicts in more than 30 countries around the world. It is estimated that more than 2 million children have died over the last decade as a direct result of armed conflict, and at least 6 million have been seriously injured or permanently disabled. In addition, between 8,000 and 10,000 children are being killed or maimed by landmines each year.

documented, this paper will explore the experience of women enslaved in Mauritania. This overview of chattel slavery in Mauritania, complemented by a confidential IRB-approved case report, is meant to inform the care of survivors of Mauritanian chattel slavery and potentially any patient with a similar history of human rights violations.

Case Report: History

> Ms. B is a 25-year-old female from Mauritania with no prior medical care. She was referred to the Center for Refugee Health and Human Rights at the Boston Medical Center by her attorney for medical and psychological evaluation in preparation for her asylum application. Ms. B was born into slavery, as were previous generations of her family. She was not allowed to own any possessions and worked 7 days a week without any compensation. She was not allowed any form of education. During her enslavement, Ms. B experienced several forms of torture including serving as her master's concubine before entering puberty, being buried to her neck in the ground, and enduring beatings and deliberate burnings as punishment for transgressions. Consistent with her cultural tradition, Ms. B underwent female genital mutilation at a young age. She escaped from Mauritania by stowing away on a ship and sought asylum in the USA.

Chattel Slavery in Mauritania

Mauritania's 1961 Constitution and this nation's affirmation of numerous human rights documents[1] collectively condemn the practice of slavery. Nevertheless, ethnic, religious, social, and gender discrimination have fostered the persistence and discouraged the practical abolition of economic and physical slavery in Mauritania [6, 7]. Economic slavery can be thought of as bondage maintained by slaves' inability to be financially self-sufficient, while physical slavery can be understood as bondage perpetuated by physical forces that prohibit freedom of movement and association [7]. Economic and physical slavery themselves breed psychological slavery. The High Commissioner concurs, suggesting, "even when abolished, slavery leaves traces. It can persist as a state of mind—among its victims and their descendants and among the inheritors of those who practiced it—long after it has formally disappeared" [1]. A continued mental state of slavery in Mauritania has justified the maintenance of barriers preventing slaves from receiving education, training in marketable skills, and the ability to learn about and advocate for their political rights. The normalcy of slavery has also been upheld by propaganda asserting that serving masters is a religious duty of slaves that if shirked, results in eternal damnation. Moreover, many Mauritanian slaves are psychologically secure in their slavery [8], given that freedom in Mauritania may lead to poverty, job discrimination, and relegation to prostitution. Escaping slavery may also result in torture if caught, alienation in a society organized by extended

[1] Mauritania ratified the League of Nation's *Slavery Convention of 1926*, the *Supplementary Convention on the Abolition of Slavery, the Slave Trade, and Institutions and Practices Similar to Slavery* in 1986, the International Labor Organization's *Forced Labor Convention* in 1961, and the ILO's *Abolition of Forced Labor* in 1997. Being a member of the United Nations, Mauritania at least nominally proclaims and promises to promote the human rights principles embedded in the 1948 *Universal Declaration of Human Rights*. Furthermore, Mauritania acceded the *Convention on the Elimination of All Forms of Discrimination Against Women* in 2001, and both the *International Covenant on Civil and Political Rights* and the *International Covenant on Economic, Social, and Cultural Rights* in 2005.

families and ruled by the *bizan* caste, and damnation [6, 8]. Only approximately 100 *abid* escape from bondage in Mauritania per year [9].

Although female *abid* are in no way a homogeneous population, nearly all female *abid* share an ambivalent position in Mauritanian society [10]. Female *abid* are preferred over their male counterparts because of their greater versatility in the division of labor [4]. Though female *abid* can make intimate alliances with both free men and women in *bizan* society, they are continually rejected and deprived of their human rights. Historically, the primary role of female *abid* has been that of a domestic worker or tent slave woman. Domestic labor persists, yet to what extent is not clearly understood. While Ruf reports that it is mainly an outdated phenomenon in Mauritania [10], a former Mauritanian slave-owner recently affirmed the contrary [11].

Tent slave women are meant to unburden the women of the house of all inconsequential tasks [10]. The two most dishonorable chores that any woman in *bizan* society can do is pounding millet and fetching water, both of which are only done by female *abid* or machines. True to their ambivalent status in *bizan* society, tent slave women also have the privilege of entering their masters' homes—normally off-limits for slaves—by serving as their masters' concubines or their mistresses' wet nurses. Both forms of alliance allow female *abid* to achieve social ascension and individual respect. Marriage with a *bizan* man provides a concubine with manumission (though she is never accepted by the world of *bizan* women) and their children share their father's status. Engaging in the institution of milk kinship enables slave women to experience a sense of belonging, enjoy social motherhood, and obtain "the right to locate themselves within the core locality of femininity," i.e. the master's tent [10].

As economics and technology change Mauritanian society, slave practices have not ended. *Bizan* women simply adapt these practices in order to maintain their superiority over *abid*. Since the severe drought of 1969, the destruction of Mauritania's economy has left many *bizan* impoverished. Poverty has prevented many free women from living up to the society's conceptions of ideal femininity: obesity and inactivity. If free women must be active in the home, they can redraw cultural boundaries in order to sustain the differences between *abid* and *bizan* by having slave women perform the most dishonorable jobs in the home. Free women can also deprive female *abid* of respect by disregarding the traditional practice of a strict gender division of labor and thus having slave women engage in men's work outside the camp: herding and cultivation. While pastoral work is no longer common because of sedentarization and the loss of animal capital, women have nearly supplanted men in the fields because of male migration to the cities. When slave women engage in cultivation, they are denied their femininity. "The locus of this battle is both the nature and the location of work" [10].

Psychological Sequelae of Mauritanian Chattel Slavery

> Ms. B's review of symptoms included pelvic pain, headaches, body pain, nightmares, sleep disturbance, flashbacks, and exaggerated startle response. She also suffered from hypervigilance, particularly that "the Master" would find her and return her to his plantation where she would be punished for running away. Ms. B reported witnessing public beatings on her owner's plantation in which recaptured slaves were killed.

The emotional dimensions of slavery have received little historical attention [12]. Only one social scientist, Orlando Patterson, has attempted to investigate and publish his findings on the psychology of Mauritanian slavery. Patterson defines a slave as a "socially dead person," alienated from all 'rights' or claims of birth, and not belonging in his/her own right to any legitimate social order [4]. Furthermore, female *abid* have no independent social existence, as evidenced in their characteristic inability to speak in their own words or consider themselves apart from their masters [10]. Their social identity is bound to their roles within their masters' families which, paradoxically, is the place where they are dehumanized, de-socialized, and de-gendered [10]. Thus, in the midst of serving their masters and mistresses, female *abid* face a terrible and confusing mixture of affection and inclusion, as well as deprivation, rejection, and exclusion.

Made to think that they are helpless children, female *abid* tend to believe that they require their masters to exist. Moreover, prevented from interacting with or marrying people of their own choosing while enslaved, female *abid* learn to depend on their masters' family for affection. This emotional attachment persists even after slaves escape. Also, told to believe that they were placed on earth to serve their masters, female *abid* look to their masters for salvation. Thus, some slaves who have escaped have willingly returned to their masters in order to again receive their master's provision and protection in Mauritania's hostile and discriminatory society [13]. *Bizan*, on the other hand, cannot function without their slaves and would never consider voluntarily freeing them; such an action would be a great dishonor [11]. As a result, a tremendous co-dependency exists between slave women and their masters. Bales describes this co-dependency as an "insidious mutual dependence that is remarkably difficult for slaveholder as well as slave to break out of" [14].

A great measure of personal resiliency is required to continually face the assertion of otherness when living in close proximity to mistresses and engaging in de-socialized and de-gendered slave labor. Not surprisingly, the highest aim of slave women is to become what they are not: they desire to adopt the key characteristics of a noblewoman, including inactivity, obesity, and the practice of a few, most often collectively performed tasks requiring only minimal physical input [10]. Social and behavioral scientists have shown that people having relatively less political power, cultural assets, social assets, honorific status, and human resources than others around them suffer from more physical and mental health problems than people residing in more equitable environments where such gaps in control and resources do not exist [15]. Slave women, therefore, are prone to developing a sense of otherness and shame from always seeing, yet never enjoying, the benefits of being a feminine noblewoman.

Female *abid* are also prone to unique perceptions of powerlessness because their social identity is determined by their current husband. If a slave woman is forced to divorce and marry another freeborn man, she must leave her children and friends from the previous marriage behind. Frequent rapes [16] and forced prioritization of masters' children over slaves' children enhance chattel slaves' feelings of powerlessness and shame [3]. Female *abid*'s role as sexual pleasure objects and their precarious connection with blood and self-selected community ties underscore how *bizan* society views these females as socially dead. Indeed, they are considered ultimate human tools—imprintable and versatile non-persons [4].

Not surprisingly, female *abid* are likely to suffer from a slave mentality. If, when, and to what extent this slave mentality develops, will impact slave women's mental health, their resulting psychological needs, and the timing of their healing process. The longer a female *abid* internalizes her slave identity, the more extended and intensive her treatment will likely be. The slave mentality is marked by three main characteristics: disconnection, disempowerment, and self-blaming. An overwhelming sense of otherness, namelessness, and invisibility can lead to disconnection from others. Endless personal violation and violence intended to maintain a reduced sense of physical and mental integrity are likely to promote feelings of powerlessness. The experience of chronic dishonor, as well as an inability to be like mistresses, care for children, and change their situations, can lead to the projection of shame and outward demonstrations of self-hatred.

Depending on the severity of a slave woman's "servile personality" [4], she may experience corresponding levels of mental exhaustion and depressed ego function [17, 18] that overshadow her will to defend her own interests within bondage, and if freed, to live her freedom [11]. Also depending upon the presence and/or nature of her slave mentality, she may lack interest in becoming aware of her social and religious deprivations and may not feel responsible for her own fate [10]. Therefore, individuals may struggle with the internalized messages of slavery for years after formal manumission or escape. Working through these messages may be even more difficult when it takes place outside of a former slave woman's country, culture, and community, since her relocation-related feelings of alienation and non-efficacy may impede her healing process. Survivors' ability and/or desire to recognize and accept their personhood and desert of human rights may be the final arbiter in determining whether they will win their psychological struggle to become like the *bizan*, the free people. Health professionals when given the opportunity, can assist female survivors of Mauritanian chattel slavery in finding strength within the human spirit to reclaim their social identity, femininity, and personhood—perhaps for the first time.

The situation presented by patients with a history of chattel slavery, though terrible, is not hopeless. Indeed, the institution of slavery is "oppressively weighted against the slave" [4], yet many women, while still in slavery, constantly struggle with the "weapons of the mind and soul" to "minimize the burden of (their) exploitation and enhance the regularity and predictability of (their) existence" [4]. Thus, female *abid* must not be dismissed as "passive victims of their brutal masters" [10] or "powerless, isolated, and degraded" persons lacking spirit, self-interest, and will [4].

Survivors of Mauritanian Chattel Slavery in USA

Roughly 3,000–4,000 Mauritanian refugees and asylees reside in the USA [9]. In 2006, 88 Mauritanian refugees arrived in the USA; 218 Mauritanians gained asylum defensively and 12 were granted asylum affirmatively [19]. Unfortunately, there is no reliable data on the number of Mauritanian survivors of chattel slavery in the USA. Female survivors of Mauritanian chattel slavery who arrive in the USA are likely to have characteristics that distinguish them from the female *abid* discussed earlier that may affect their health care needs. First, they are more likely to have suffered physical violence from their masters [8]. Second, they may have a stronger sense of identity and a greater awareness of or longing for freedom. Third, they may have had an opportunity to learn about their political and civil rights in Mauritania or develop critical thinking skills from their masters' children who themselves questioned slavery and thus secretly educated their and/or their parents' slaves. Lastly, they are likely to have exceptional inner strength, evidenced by their ability to leave at least

part of their families behind, escape from their masters, and journey to the USA.

These differences should help survivors of Mauritanian chattel slavery progress along their healing journey. Though the occurrence of physical violence is tragic, it may help patients recognize that their former *bizan* are not perfect and superior beneficent providers, saviors, or gods, and rather are fallible human beings. Histories of physical violence may also ignite in survivors an intrinsic need to defend themselves. This desire to protect self can be harnessed in the recovery process. The self-awareness and development of ego functions made possible by the other three distinctions between *abid* in Mauritania and survivors of Mauritanian chattel slavery in the USA should all aid in the humanizing, connecting, and empowering goals of mental health and medical treatment.

Proposed Framework for Mental Health and Medical Treatment

No formal research has been published on caring for female survivors of chattel slavery. Nevertheless, western psychology and psychiatry offer three helpful concepts that may facilitate health professionals in serving this and similar patient populations. The first concept is complex posttraumatic stress disorder (PTSD) (Table 2), characterized by systematic and pathological changes in a person's affect regulation, consciousness, self-perception, identity, perception of her perpetrator, relations to others, and systems of meaning [17]. Herman describes complex PTSD resulting from chronic trauma as "classic post-traumatic syndrome" coupled with "profound alterations in (patients') relations with God, with other people, and with themselves" [18]. Complex PTSD involves the fusing of "protracted depression" and "chronic hyperarousal and intrusive symptoms" to create "the 'survivor triad' of insomnia, nightmares, and psychosomatic complaints" [18]—all of which have been reported by survivors of slavery [3]. Complex PTSD further entails the merging of depressive symptoms with chronic trauma's sequelae to form destructive couplets more pathological than the sum of their parts: dissociation with concentration difficulties, apathy and helplessness with paralysis of initiative, isolation with disruption in attachment, depressive guilt with debased self-image, and hopelessness with loss of faith [18]. Patients with complex PTSD also project their intense anger with their perpetrators onto themselves. There are no data specifically on the prevalence of complex PTSD in a population of former slaves, however in an urban population of adult refugees, 9% have experienced PTSD symptoms [20].

Table 2 Complex PTSD symptoms [3, 4, 7, 8, 17, 18]

Alterations in emotional regulation
Including: persistent sadness, suicidal thoughts, explosive anger, anger projected towards self
Alterations in consciousness
Including: chronic hyperarousal, forgetting and/or reliving traumatic events, dissociation, concentration difficulties, insomnia, nightmares, psychosomatic complaints
Alterations in self-perception and identity
Including: apathy, helplessness, shame, guilt, stigma, reduced self-efficacy, sense of being completely different than other human beings
Alterations in the perception of the perpetrator
Including: attributing total power to the perpetrator, preoccupation over the relationship with the perpetrator, preoccupation with revenge
Alterations in relations with others
Including: isolation, distrust, repeated search for rescuer or new master
Alterations in one's system of meanings
Including: loss of sustaining faith, sense of hopelessness and despair

Psychology's second applicable tool is that of mental defeat, which results from multiple, severe, prolonged, and inescapable aversive non-physical attacks done in the context of totalitarian control. Mental defeat can be understood as a perceived loss of autonomy, choice, and free will, as well as the belief that one's identity cannot be maintained [17]. Not all who experience loss of autonomy develop mental defeat, suggesting that mental defeat is an intermediary step between loss of autonomy and mental death. Furthermore, mental defeat may be predictive of PTSD incidence and severity [17].

The third concept is that of mental death, which is a form of complex PTSD that arises from threats to psychological, rather than physical, integrity [17]. Mental death is characterized by a loss of identity possessed before experiencing interpersonal trauma within the context of totalitarian control. It is associated with guilt and shame; distrust and alienation from others; feelings of ineffectiveness, inability to actively cope, loss of self-efficacy, and loss of autonomy; loss of core beliefs and values; and a sense of being permanently damaged.

While the concepts of complex PTSD, mental defeat, and mental death all have similarities with the concept of "slave mentality," they may not be fully adequate in instructing health professionals on how to best care for survivors of slavery. Used in understanding sequelae of organized political torture, complex PTSD, mental defeat, and mental death all assume a pre-trauma identity, as well as alterations or losses to that identity. Women who have been enslaved since birth, however, may have never had such an identity or conceived of a world that does not include their trauma. The

absence of pre-trauma histories, as well as the presence of complex, chronic trauma histories may make conducting psychological assessments and creating treatment plans more difficult. Chronic trauma and its presence during numerous stages of a person's life make reconstruction of trauma history very complicated [18]. Furthermore, disorganized and fragmented memories—characteristic of chronic trauma—are difficult to process [17]. The confusing nature of relationships in chattel slavery and the long-term nature of the trauma endured by survivors of chattel slavery will likely slow their progress. Conversely, survivors of chattel slavery cannot rely on pre-trauma life experiences to guide their current and future behaviors [17].

Mental Health Treatment

Since "real liberation takes place in the mind" [14], survivors of chattel slavery coming to the USA may benefit from mental health care. At best, former slaves who are not ready to hear about or embrace freedom simply laugh at the concept [11]. At worst, they attempt suicide when slavery—the major psychological building block of their lives—is challenged [14]. Therefore, it is important that health professionals recognize the former slave status via an appropriate patient history and then only treat those who desire recovery and freedom. Herman supports this assertion, while Gray et al. [13, 18] suggest that in situations of limited resources, treatment should be further narrowed to those patients who are most at risk for persistent distress.

Although there are no data on effective treatment methods specifically for former chattel slaves, treatment can be initially modeled after protocols for survivors of other forms of trauma. Treatment should be culturally sensitive [21], early, active, empowering, supportive, multidisciplinary, and structured [22]. Moreover, no treatment regimen has been developed specifically for complex PTSD, though it is known that exposure therapy is not helpful [17]. In general, therapists should address survivors' disempowerment, disconnection, co-dependency, self-blaming, and PTSD symptoms. Offering moral solidarity, rather than clinical neutrality, will be especially necessary in caring for survivors of chattel slavery as they learn to embrace and explore their freedom for the first time [18]. Medications have been helpful as adjunct therapy for treating complex PTSD symptoms and major depressive disorder.

Survivors of chattel slavery will likely require a combination of cognitive, interpersonal, and social strategies over the course of a long period of time [18]. Benchmarks for the effectiveness of treatment focused on empowerment should include indications of identity formation, self-efficacy, self-worth, the ability to go to school or work, the aptitude to care for oneself and one's family, the capability to act on dreams and aspirations, and the capacity to formulate and express one's own sense of spirituality. Recovery surrounding relational issues is marked by the lessening of fear, shame, guilt, self-hatred, anxiety, and sense of contamination; the experience of respect and honor; the ability to foster, develop, and maintain safe and meaningful alliances and relationships [23]; and the capacity to enjoy healthy forms of intimacy. Other benchmarks for healing include the absence of or ability to control PTSD symptoms, the formation of a coherent narrative of trauma linked with emotions, and the construction of a coherent system of meaning and belief that encompasses the trauma story [18, 21].

A sense of temporal and eternal safety should be established early in treatment. Temporal safety enables survivors of chattel slavery to feel less vulnerable, to gain a sense of predictability, and to develop self-efficacy and trust. Learning how to control their bodies and their environments as much as possible will aid survivors in experiencing temporal safety [18]. Because survivors' bodies were literally owned by others, working towards ordering their actions may be especially difficult. A sense of eternal safety will also be important if survivors of chattel slavery are to embrace and learn to live their freedom, as fear of damnation resulting from fleeing bondage will likely haunt these patients. Imams can be instrumental in correcting misconstrued religious beliefs that were used to justify slavery. Helping patients adopt a less hierarchical version of Islam will also be helpful in combating remnants of pro-slavery thinking. Therapists should be aware that female survivors of chattel slavery may be hesitant in speaking with Imams because women in Mauritania are forbidden to receive religious instruction.

Helping survivors become aware of their humanity will also be necessary for healing to occur. Mental health professionals should assist patients in recognizing their emotions, sense of morality, talents, and ambitions. Mourning loss will be an important stage for therapists and survivors of chattel slavery to explore. Like mental health professionals caring for survivors of child abuse, therapists working with survivors of chattel slavery will need to walk alongside patients as they mourn what was never theirs to lose [18]. Patients' anger at the system of slavery in general and masters in particular can promote recovery, as long as this anger can be channeled in a constructive way. The desire for freedom, as well as for the equality, free choice, paid labor, and ability to act on preferences that it can bring, will be a great ally in the recovery of chattel slavery survivors.[2]

[2] "'When a slave runs away,' says Nasser, 'he's losing his roots. Slavery is his reality.' In such a context, 'it would take a person of enormous energy, with a built-in quest to find a new life, to stand up and walk away,' says Robert Pugh, a former US ambassador to Mauritania" (Skinner 43).

Assisting survivors in becoming self-sufficient will be very important in furthering their recovery. In individual therapy, mental health professionals can draw out patients' unique coping mechanisms and have them recount the various ways in which they resisted the system of slavery while still in bondage. Encouraging survivors of chattel slavery to take part in education, life skills, and occupational training programs will further their sense of self-efficacy. Vocational rehabilitation specialists will also be important in helping survivors of chattel slavery formulate and work towards personal goals and help them develop autonomy. Aiding survivors of chattel slavery learn life-skills, such as planning a daily schedule, will be especially important since former slaves have reported difficulties in getting out of bed and performing daily activities without their masters to direct their steps [11].

Aiding survivors in developing dialectical thinking is important, as this will help them adopt new beliefs and attitudes in a healthy manner [17]. For survivors of chattel slavery, slavery was and to some extent probably still remains a normal part of life. Challenging vestiges of patients' old belief systems and having them reconstruct worldviews that affirm the injustice of slavery will be painful. This is the irony of caring for survivors of chattel slavery—therapy may be the cause of just as much, or even more, cognitive upheaval than the original trauma [18]. Literacy and education classes will help survivors of chattel slavery cope with the cognitive dissonance that inevitably results from realizing that they once affirmed a system of life to which they can no longer ascribe.

In situations where race is used to justify chattel slavery, further cognitive restructuring should be done to address the assertion that dark skin indicates impurity [11]. Also, therapists should help survivors of chattel slavery in recalling events during which they witnessed the fallibility of their masters [11]. By simultaneously helping survivors discover their own identities as human beings and working with them to understand that they are not inferior, but rather are of equal standing with their masters, therapists can assist survivors of chattel slavery in countering feelings of otherness, degradation, and disenfranchisement [18, 21]. Once survivors accept that their former masters have no power over them, they can develop the ability to live apart from and uphold different beliefs than their masters. Such will further the process of empowerment and connection necessary for recovery [18].

Developing the ability to establish beneficial, appropriate connections with others will also be an important goal for survivors of chattel slavery. Individual therapy can assist survivors in learning how to trust and be known by another within a supportive, collaborative atmosphere. Therapists should be aware that the risk of transference and counter-transference is especially great because of the tremendous power differential between mental health professionals and survivors of chattel slavery [18]. A survivor may come to think of her therapist as her new mistress [22], making it difficult for the patient to enter a collaborative physician–patient relationship [18]. Conversely, mental health professionals may struggle with wanting to fulfill the savior-like role imposed on them by their patients [3, 24]. Because of possible rape histories and Mauritania's fundamentalist culture, female therapists should work with female survivors of chattel slavery whenever possible.

In addition to individual therapy, group therapy may provide survivors of chattel slavery with a space to engage in psychoeducation, open group discussion, cognitive restructuring, maladaptive interpersonal relationship work, and problem solving. Group therapy has been shown to reduce some childhood trauma symptoms [25] and help trauma survivors develop and maintain safe relationships, as well as explore how past trauma affects current relationships [23]. Group therapy may prove to be especially useful in helping survivors of chattel slavery combat common fears, including anxiety about the responsibility of living on one's own and the belief of damnation resulting from fleeing bondage. Patients may also fear that their masters will find them in the USA and punish them for escaping. Moreover, survivors of chattel slavery may suffer from knowing that without formal papers of manumission, they will never be considered free in their home country. Guilt from abandoning their own and their masters' families, the slave mentality, psychological symptoms, fatalism, expectations of a short life-expectancy and the consequent devaluing of future-focused thinking [22], and negative attitudes towards mental health treatment [22] are also significant barriers to healing. It has been shown that those with mental defeat or mental death have more severe PTSD and are more treatment-resistant [17]. Furthermore, those with mental death are reportedly less likely to have optimistic expectations about therapy and have more difficulty developing effective working relationships with therapists [17].

Social support systems, whether formal or informal, will be very important in helping survivors of chattel slavery gain a sense of connection and solidarity with others as equals. Having survivors of chattel slavery become involved in their communities and take part in social obligations—such as *zakat*[3]—will help them join a moral and social space in which their humanity and dignity are affirmed. Holding some form of public celebration in which survivors can declare their freedom and personal identity may serve as an important milestone for some patients.

[3] *Zakat* is tithing for the poor prescribed by Islam (Ruf 263).

During therapy sessions, language barriers, cross-cultural issues, transference, counter-transference, vicarious traumatization, patients' dishonesty [24], and the painful cognitive restructuring embedded in the therapy process itself will hinder recovery. The stressful asylum-seeking process, discrimination of all kinds, and re-victimization in the form of unhealthy relationships at best or new slavery at worst will delay the establishment of safety in the USA. In addition to having compromised abilities to defend their own interests, survivors of chattel slavery may not resist new slavery in the USA because it typically involves work with which they are familiar and may enjoy—domestic services, agricultural work, street peddling, and restaurant services. An absence of standard of care guidelines for this patient population, as well as resource limitations and difficulties in finding safe, affordable housing for slavery survivors will all likely impede the recovery process [3].

One of the greatest challenges that face both survivors of chattel slavery and their health professionals is residual religious-based messages yoked to the concept of slavery. Patterson asserts that masters' authority "rests on the control of those private and public symbols and ritual processes that induce (and seduce) people to obey because they feel satisfied and dutiful when they do so" [4]. In Mauritania, masters exploit Islam to justify and maintain current oppressive practices [6, 12]. *Bizan* and religious leaders teach *abid* that slavery is a duty that must be fulfilled in order to please Allah. Such indoctrination, evidenced in the common Mauritanian phrase, "the way to heaven is underneath the sole of your master's foot" [6], entangles masters with the concepts of mediator, savior, and God. Masters are not only mediators between secluded slaves and the world outside their masters' camps, but between earth and heaven. Thus, a former Mauritanian slave woman spoke [6]:

> We hear of abolition, but for most slaves it does not mean much. It is hard to ignore what they have been told all their lives, that without their master they cannot survive, that only he can ennoble them, give meaning to their life and lead them to heaven. They believe this; so how can they also believe that they must escape the situation that promises to give them so much? The slave lives in perpetual awe of his masters and is not aware of any other way of life.

The exploitation of faith and concepts of salvation are particularly dangerous for the minds and souls of slave women. Being female, they will have the least opportunity to learn what Islam truly asserts regarding God, their relationship with God, slavery, and the actual duties called for by Islam [10]. Ignorance, the need to hope for a better future in order to endure present suffering [8], and the willingness of many men and women throughout human history to preserve eternal rather than temporal life according to their belief systems [26] may be the most difficult barriers to successfully caring for the mental health of women with a history of chattel slavery.

Medical Treatment

Case Report: Physical Exam

> On initial visit, Ms. B's affect was numb and flat. She displayed little emotion, even when describing horrific events. Ms. B was unable to give a chronological history of life events, had no sense of linear time, and lacked spontaneous speech. Physical examination was difficult because of her fear and was terminated early during the abdominal/external pelvic exam due to a terror reaction. Ms. B's skin was significant for multiple scars and healed burn marks. The limited abdominal exam revealed a tender suprapubic mass and follow-up transabdominal pelvic ultrasound revealed an adnexal cystic mass. Ms. B was referred for exploratory laparoscopy. Prior to the procedure, she was given a tour of the operating suite and a detailed explanation of what she should expect on the day of surgery. During the procedure, she had a pelvic examination and pap smear under general anesthesia. Ms. B's primary care physician arranged to be in the preoperative area to give reassurance and in the recovery suite to provide support and monitor for posttraumatic stress disorder symptoms.

Providing medical care to women with a history of complex trauma, including chattel slavery, requires the establishment of trust and rapport through empathy, respect, patience, and active listening. Former slaves seeking health care in the USA may have never received any prior medical care and may be unfamiliar with bio-western medicine. Undergoing physical examinations, phlebotomy, routine screenings, and vaccinations may be foreign. In addition, survivors of chattel slavery may have difficulty answering questions routinely used to gather patient histories. Echolalia, monotonic speech, flat affect, paucity of rich language structures, difficulty expressing complex thought, and ignorance of age may be due to prolonged isolation, lack of social stimulation, and little practice with two-way dialogue. Medical providers should be familiar with and evaluate the former slave for physical sequelae of torture. Signs and symptoms can include sexual trauma, scars, musculoskeletal injuries, and head trauma.[4]

[4] The online course, *Caring for Torture Survivors*, provides an extended discussion (with illustrations) on mental and physical sequelae of torture, as well as how mental health, medical, and oral health professionals should approach such sequelae (www.bcrhhr.org).

The environment for the clinical encounter and physical exam should be made as comfortable and safe as possible. Selecting appropriate décor and limiting the number of medical instruments easily visible help in creating a safe atmosphere. Using trained, same-sex interpreters in caring for traumatized women who do not speak English is recommended. Health professionals at the Boston Center for Refugee Health and Human Rights have found telephone interpreter services very helpful. Survivors of chattel slavery are accustomed to lacking control over internal and external loci. Consequently, physicians may need to provide patients with encouragement in exercising control within clinical encounters. For instance, physicians should give patients explicit permission to terminate either the interview or examination whenever the patient feels uncomfortable. Moreover, patience and understanding is necessary as survivors learn to think about and perform multi-step tasks within and outside the clinic (i.e. giving voluntary informed understanding consent and medical adherence).

In addition to addressing the medical needs of patients, clinicians can serve an important function in the asylum seeking process, as affidavits documenting physical and psychological sequelae are often helpful in supporting asylum claims. Physicians who care for survivors of slavery will need to work closely with legal professionals, as the asylum-seeking process allows survivors to begin the acculturation process and assume the responsibilities and privileges corresponding with asylum. Specifically, gaining asylum gives survivors of chattel slavery legal and social standing, as well as the ability to work, which will be instrumental in satisfying patients' needs for empowerment and connection.

Case Report: Follow-Up

> Ms. B has made great strides despite the inadequacies of the medical community's understanding of how to effectively address her health needs. She has gained a full range of emotions and speech, her mental health symptoms have abated, and she continues to receive supportive care. With the help of an affidavit written on her behalf by her physician at the Boston Center for Refugee Health and Human Rights, Ms. B was granted asylum. She has begun to learn English and is now living in safe housing with two family members.

Agenda for Further Research

The above recommendations on caring for female survivors of chattel slavery have not been confirmed by research. Instead, these paradigms were developed from the limited information known about the socio-historical context of chattel slavery in Mauritania, as well as from best practice guidelines for patients with histories of other types of torture and trauma. Application and evaluation of these recommendations is necessary in order to establish sensitive and competent protocols that support the mental health and medical needs of all female survivors of chattel slavery coming to the USA.

Research should be directed at identifying the psychological symptoms and needs of chattel slavery survivors. Such research would ideally be used to accept or reject the possible relationships among mental defeat, mental death, slave mentality, and chronic PTSD proposed in this paper. Once the mental sequelae of chattel slavery is better understood, prospective studies should be performed in order to determine how to best care for the mental health of chattel slavery survivors. Data collection could include survivors' beliefs about mental health and medical treatment, as well as patient responses to various treatment approaches in various clinical settings.

Research should be preceded by consultations with institutions and professionals who have worked with survivors of chattel slavery.[5] Such consultations can provide information on how experienced professionals understand the psychological impact of chattel slavery from working with people with a history of such trauma. Consultations may also shed light on how experts have treated or supported survivors of slavery, what they have found to work and not work, and what questions they still have.

Because few survivors of chattel slavery have been identified in the USA, research endeavors may need to go beyond female survivors of chattel slavery to include male survivors, and perhaps men and women with histories of other forms of slavery. Broadening the subject pool may be necessary to ensure valid and generalizable data. Whether and when formal research should be done with survivors of slavery is a question that is beyond the scope of this paper. Such a decision involves a myriad of ethical dilemmas—typically surrounding the principles of beneficence and autonomy—that must be wrestled with in a serious

[5] Experts that may be helpful can be found at the Center for Refugee Health and Human Rights (http://www.bcrhhr.org), Iabolish (http://www.iabolish.org), Free the Slaves (http://www.freetheslaves.net), the Bellevue/NYU Program for Survivors of Torture (http://survivorsoftorture.org/survivors), the Department of Sociology and Anthropology at Georgetown University (http://www3.georgetown.edu/departments/sociology/about), the Program for Survivors of Torture and Severe Trauma: Center for Multi-Cultural Studies (http://ncttp.westside.com/wsContent/default.view?_pagename=VA-Program+for+Survivors), and the UUA Holdeen India Program (http://archive.uua.org/international/holdeen/panditbio.html).

manner.[6] If experts working with and knowledgeable about survivors of slavery agree that this population should be involved in research to help therapists learn how to best care for them, a retrospective study could be conducted to formally understand how survivors of slavery have responded to various interventions thus far. In addition, prospective research directed at determining which forms of mental health interventions are most effective could be performed. All the data collected from consultations, interviews, and research can be used to create best practice guidelines. These guidelines can be disseminated via the internet—discussed and improved by experts over time—and hopefully will aid health professionals to sensitively and competently care for survivors of chattel slavery worldwide.

Conclusion

Caring for former Mauritanian chattel slaves presents unique challenges. The trauma of chattel slavery, and particularly the mental health consequences of such trauma, will influence all encounters this patient population has with the medical community. Research is needed to sensitively and competently inform the best methods of care for the mental and physical health needs of female survivors of chattel slavery. Even before research is completed, health care professionals should continue to care for patients surviving slavery, bearing in mind the type of slavery endured, the gender of the patient, and the socio-political-cultural-religious context in which bondage was experienced.

References

1. Office of the UN High Commissioner for Human Rights: Fact Sheet No. 14: Contemporary Forms of Slavery. Geneva: United Nations; 1991.
2. Dodson H. Slavery in the Twenty-First century. UN Chronicle Online Edition 2005; vol XLII, November 5.
3. Bales K. New slavery: a reference handbook. 2nd ed. Santa Barbara, CA: Contemporary World Issues Series; 2004.
4. Patterson O. Slavery and social death: a comparative study. Cambridge, MA: Harvard U Press; 1982.
5. Mohamed A. African country to jail slave owners. The Boston Globe 2007. http://www.boston.com/news/world/africa/articles/2007/08/10/african_country_to_jail_slave_owners/.
6. Fleischman J. Mauritania's campaign of terror. Africa, NY: Human Rights Watch; 1994.
7. Working Group on Contemporary Forms of Slavery: Contemporary Forms of Slavery, Report on 30th Session. GE.05-14789 (E) 150705; 2005.
8. Bales K. Mauritania: old times there are not forgotten In: Disposable people: new slavery in the global economy. Berkeley: U of CA Press; 1997. p. 80–120.
9. Sage J. Personal correspondence; 11/17/05.
10. Ruf UP. Ending slavery: hierchary, dependency, and gender in central Mauritania. London: Transaction Publishers; 1999.
11. Nasser A. Lecture: a slave owner's rebellion; 11/17/05.
12. Wyatt-Brown B. Psychology. In: Finkelman P and Miller JC, editors. Macmillan encyclopedia of world slavery. London: Prentice Hall International; 1998:2 747–8.
13. Gray M, Litz BT. Behavioral interventions for recent trauma: empirically informed practice guidelines. Behav Modif. 2005; 29(1):189–215.
14. Bales K. The social psychology of modern slavery. Sci Am. 2002;286(4):80–8.
15. Kawachi I. Income inequality and health. In: Berkman LF, Kawachi I, editors. Social epidemiology. NY: Oxford U Press; 2000. p. 76–94.
16. Iabolish. Breaking free from the bonds of slavery: an interview with Maalouma Messaoud, slavery survivor and Mauritanian abolitionist. http://www.iabolish.com/act/abol/profile/maalouma.htm (2005).
17. Ebert A, Dyck MJ. The experience of mental death: the core features of complex posttraumatic stress disorder. Clin Psychol Rev. 2004;24:617–35.
18. Herman JL. Trauma and recovery: the aftermath of violence—from domestic abuse to political terror. NY: Basic Books; 1992.
19. US Department of Homeland Security: Yearbook of Immigration Statistics: Refugees and Asylees; 2006. http://www.dhs.gov/ximgtn/statistics/publications/YrBk06RA.shtm.
20. Crosby SS, et al. Emergence flashback in a patient with posttraumatic stress disorder. Gen Hosp Psychiatry. 2007;29:169–71.
21. Gusman FD, et al. Multicultural development approach for treating trauma. In: Mareslla AJ, et al., editors. Ethnocultural aspects of posttraumatic stress disorder: issues, research, and clinical applications. Washington, DC: APA; 1996. p. 439–58.
22. van der Veer and van Waning. Creating a safe therapeutic sanctuary. In: Wilson JP, Drozdek B, editors. Broken spirits: the treatment of traumatized asylum seekers, refugees, war, and torture victims. NY: Brunner-Routledge; 2004. p. 187–220.
23. Harner HM. Relationships between incarcerated women: moving beyond stereotypes. J Psychosoc Nurs Ment Health Serv. 2004;42(1):39–46.
24. Physicians for Human Rights: examining asylum seekers: a health professional's guide to medical and psychological evaluations of torture. USA: PHR; 2001.
25. Kanas N. Group therapy for patients with chronic trauma-related stress disorders. Int J Group Psychother. 2005;55(1):161–5.
26. Munson R.: Intervention and reflection: basic issues in medical ethics, 7th ed; 2004. p. 124–5.

[6] For instance, patients should ideally not be given treatment that is not proven safe and effective. Safety and effectiveness, however, typically require formal research. What happens when the patient population needing treatment cannot give voluntary understanding informed consent? Is it better for therapists to determine best treatment guidelines via formal research without such a stringent form of consent, or for therapists to continue to anecdotally create such guidelines until the high bar of voluntary informed understanding consent can be reached?

How Valid is the AHRQ Patient Safety Indicator "Postoperative Physiologic and Metabolic Derangement"?

Ann M Borzecki, MD, MPH, Marisa Cevasco, MD, MPH, Qi Chen, MPH, Marlena Shin, JD, MPH, Kamal MF Itani, MD, Amy K Rosen, PhD

BACKGROUND: The Agency for Healthcare Research and Quality Patient Safety Indicator postoperative physiologic and metabolic derangement (PMD) uses ICD-9-CM codes to screen for potentially preventable acute kidney injury (AKI) requiring dialysis plus diabetes-related complications after elective surgery. Data on PMD's accuracy in identifying true events are limited. We examined the indicator's positive predictive value (PPV) in the Veterans Health Administration (VA).

STUDY DESIGN: Trained abstractors reviewed medical records of 119 PSI software-flagged PMD cases. We calculated PPVs overall and separately for renal- and diabetes-related complications. We also examined false positives to determine reasons for incorrect identification, and true positives to determine PMD-related outcomes and risk factors.

RESULTS: Overall 75 cases were true positives (PPV 63%, 95% CI 54% to 72%); 73 of 104 AKI cases were true positives (PPV 70%, 60% to79%); only 2 of 15 diabetes cases were true positives (PPV 13%, 2% to 40%). Of all false positives, 70% represented nonelective admissions and 23% had the complication present on admission. Of AKI true positives, 37% died and 26% were discharged on dialysis; 55% had chronic kidney disease (≥ stage 3) present on admission. Cardiac surgery represented the largest category of AKI-associated index procedures (30%). AKI was most commonly attributed to perioperative renal hypoperfusion (84% of true positives), followed by nephrotoxins (33%) including contrast (11%).

CONCLUSIONS: Due to its low PPV, we recommend removing diabetes complications from the indicator and focusing on AKI. PMD's PPV could be significantly improved by using present-on-admission codes, and specific to the VA, by introduction of admission status codes. Many PMD-identified cases appeared to be at high risk based on patient- and procedure-related factors. The degree to which such cases are truly preventable events requires further assessment. (J Am Coll Surg 2011; 212:968–976. © 2011 by the American College of Surgeons)

Postoperative physiologic and metabolic derangement (PMD), an administrative data-based outcome, was developed by the Agency for Healthcare Research and Quality (AHRQ), as one of 20 adult patient safety indicators (PSIs).[1] This indicator uses ICD-9-CM codes to flag cases of "specified" postoperative physiologic and metabolic derangements among elective surgical discharges.[1] Specifically, cases are flagged based on the presence of either an acute kidney injury (AKI, formerly known as acute renal failure) secondary diagnosis code and a dialysis procedure code, or a code for a diabetes complication (ketoacidosis, hyperosmolarity, or hypoglycemic coma).

Similar to other PSIs, this indicator was designed to identify inpatient complications that are clinically significant and potentially preventable. Although postoperative AKI is known to be an important complication of surgery, less is known about the included postoperative diabetes complications. The incidence of postoperative AKI varies according to the AKI definition used, type of operation, surgical urgency, and population studied.[2,3] The risk of

Disclosure Information: Nothing to disclose.

Supported by Department of Veterans Affairs Health Services Research and Development Service grant number SDR 07-002.

Received October 20, 2010; Revised December 21, 2010; Accepted January 4, 2011.

From the Center for Health Quality, Outcomes and Economic Research, Bedford VAMC, Bedford, MA (Borzecki); the Department of Health Policy and Management, Boston University School of Public Health (Borzecki, Rosen); Boston University School of Medicine (Borzecki, Itani); the Departments of Surgery, VA Boston Healthcare System (Cevasco, Itani) and Brigham and Women's Hospital (Cevasco); Center for Organization, Leadership, and Management Research, VA Boston Healthcare System (Chen, Shin, Rosen); Harvard Medical School (Itani), Boston, MA.

Correspondence address: Ann M Borzecki, MD, MPH, CHQOER, Bedford VA Hospital (152), 200 Springs Rd, Bedford, MA 01730. email: amb@bu.edu

ISSN 1072-7515/11/$36.00
doi:10.1016/j.jamcollsurg.2011.01.001

Abbreviations and Acronyms

AHRQ = Agency for Healthcare Research and Quality
AKI = acute kidney injury
EMR = electronic medical record
ESRD = end-stage renal disease
IRR = inter-rater reliability
NSQIP = National Surgical Quality Improvement Program
PMD = postoperative physiologic and metabolic derangement
PPV = positive predictive value
PSI = patient safety indicator
VA = Veterans Health Administration

postoperative AKI is highest among patients undergoing cardiac and aortic operations, with an incidence between 1% and 15% based on metabolic parameters,[2] or 1% and 5% if a requirement for dialysis is specified.[3-5] Postoperative AKI requiring dialysis is associated with more than 2-fold increased lengths of stay and in-hospital costs[6,7] and mortality rates (both in-hospital and 30-day) in excess of 60%.[4,5] Although information specific to postoperative diabetes complications is lacking, limited data show that hospital-acquired diabetes complications are rare, occurring in 0.02% of discharged patients with diabetes, but are similarly associated with excess lengths of stay, costs, and mortality.[8] Postoperative AKI prevention entails avoiding perioperative renal hypoperfusion and limiting exposure to nephrotoxic agents.[9] For diabetes, prevention includes careful blood sugar monitoring and avoidance of sliding scales as the sole method of insulin delivery, especially in patients on preoperative insulin.[10]

Given that PMD uses administrative data, the indicator (and other PSIs) was originally developed to screen for possible patient safety problem areas for local quality improvement efforts.[1] However, increasing demand for easily applied quality measures has broadened their use. PMD is now a component of a composite PSI measure that the Centers for Medicare and Medicaid Services (CMS) are tracking as part of their pay-for-reporting program.[11-13] Hospitals that underperform or fail to report these measures receive reduced Medicare payments.

The increasing adoption of PSIs as quality and patient safety measures necessitates that users understand their strengths and limitations. Like the other PSIs, PMD was developed through a rigorous process including a consensus panel of clinical experts.[14] Additionally, rates of this PSI have been found to have strong positive correlations with other postoperative PSIs,[1] and its occurrence was associated with excess in-hospital lengths of stay, costs, and deaths.[15,16] Yet, few data exist regarding this indicator's ability to detect true events as identified by medical record review. A previous Veterans Health Administration (VA) study using 2001 data compared the PMD subset of flagged cases with AKI (based on PSI version 2.1) with chart-based data from the National Surgical Quality Improvement Program (NSQIP) (n = 62) and found a positive predictive value (PPV) of 54% and a sensitivity of 44%.[17] This resulted in indicator revisions such that diagnostic codes 586 ("renal failure unspecified") and 997.5 ("urinary complications") were added to the existing 584.x code ("acute renal failure"), which slightly increased both the PPV and sensitivity, to 63% and 48%, respectively. However, a number of questions about this indicator's validity remain. Only AKI-related complications were studied; there are no data on the criterion validity of the diabetes-related complications. The small number of NSQIP cases eligible for comparison may limit the generalizability of findings, and finally, there are no data on the degree to which PMD flagged cases are associated with process of care problems and therefore potentially preventable.

In this study, we examined the PPV of this indicator in the VA. We also investigated associated risk factors and perioperative processes of care to better understand the potential preventability of this PSI.

METHODS

This was a retrospective cross-sectional study using VA administrative and electronic medical record (EMR) data from October 1, 2002 through September 30, 2007. We obtained study protocol approvals from the Bedford VA Medical Center and the VA Boston Healthcare System Institutional Review Boards.

Data sources

We used hospital discharge information, (demographics, ICD-9-CM coded diagnoses and procedures, and discharge status) from the VA's National Patient Care Database Patient Treatment File.[18] Per earlier PSI work, we eliminated nonacute care (eg, long-term care).[15,19] We accessed VA EMR data using VistAWeb, a program enabling centralized access to EMR data from all VA facilities.[20]

PMD definition

See Appendix 1 (online only) for the full PMD definition including ICD-9-CM codes.[1,21] The numerator requires a secondary diagnosis code for AKI and a dialysis procedure code, or a secondary diagnosis code for a diabetes complication of ketoacidosis, hyperosmolarity, or coma. The denominator excludes discharges when the hospitalization was nonelective, the condition was present on admission, patients had advanced chronic kidney disease (stage 5) or end-stage renal disease (ESRD) present on admission, or

dialysis occurred before or on the same day as the first operation. Additionally, patients with diabetes complications are excluded if their principal diagnosis is diabetes; patients with AKI are excluded if they have a principal diagnosis of acute myocardial infarction, cardiac arrhythmia, cardiac arrest, shock, hemorrhage, or gastrointestinal hemorrhage.

Similar to PSI 11, postoperative respiratory failure, and PSI 13, postoperative sepsis, the denominator excludes nonelective hospitalizations because these complications are less likely to be preventable "in patients admitted for non-elective surgeries, or urgent/emergent conditions."[19] However, unlike nonfederal databases that include an admission type, the VA inpatient file lacks such a variable. (Neither source includes a surgery status field; admission type is intended as a proxy for surgery status.) So, previous PSI work in the VA involved developing an algorithm to characterize admission type using diagnostic related groups, admission date and time, and principal procedure date and time.[19,22]

Study population

Hospital sampling

The sampling method and final sample are described in detail elsewhere.[22] Briefly, we applied the PSI software (v. 3.1a) to the VA inpatient database to obtain individual PSI counts and composite scores (a combined measure that includes 11 PSIs).[23,24] Starting with 158 acute care hospitals, we selected a final sample of 28 hospitals, based on PSI counts, scores, and geographic location, which was intended to reflect the diversity of VA hospitals. The final sample included hospitals from 18 of the 21 VA regional health care networks.

Case identification

We identified the eligible pool of discharges based on presumed elective admission status using previously developed criteria,[19] and randomly selected 4 flagged cases of PMD from each sample hospital, yielding 112 cases. This total was chosen to ensure reasonably narrow confidence intervals (10% to 20%) based on power calculations using previously reported PPVs of similar indicators.[25] Due to relatively few cases of diabetes complications (n = 8), we opted to include all available flagged diabetes cases from our hospital sample, for a total of 119 cases (104 cases of AKI and 15 cases of diabetes complications).

Medical record abstraction

Two trained nurse abstractors conducted EMR reviews using a standardized data abstraction instrument and guidelines adapted from AHRQ-developed preliminary tools.[26] This instrument included initial questions about demographics and ascertainment of the event; if a case was deemed a false positive, abstraction ceased; for true positives, additional information was abstracted on risk factors, evaluation, management of the event, and patient outcomes.

To ensure consistency of abstracted information, we examined inter-rater reliability (IRR). The nurses independently abstracted the same records in groups of 5 until they achieved ≥90% agreement across all questions; thereafter they abstracted different records. Study physicians (AB) reviewed questions on which nurses disagreed, with resulting instrument revisions and/or guideline clarifications as appropriate. Study physicians also reviewed cases for clarification as required throughout the abstraction process. Another IRR assessment was performed on 5 charts toward the end of the abstraction process to check for drift in abstractor reliability. Additional details of the abstraction process are available elsewhere.[22]

Nurses agreed 100% on all questions pertaining to ascertainment of the event (identification of cases as true positives or false positives) on all rounds of IRR testing (kappa = 1.0). However, we required 2 early rounds of IRR assessment to achieve the prespecified observed agreement threshold. Observed agreement on an initial round of 5 charts was 85% (n = 60 questions), mainly due to disagreements with respect to the most likely causes of AKI and how to answer procedure-related questions for patients who had multiple operations before initiating dialysis. These questions were made more explicit, and observed agreement was 98% on the next IRR assessment round. Agreement on late IRR testing was 94%.

Analyses

We classified cases as true positives or false positives based on abstracted data, and analyzed AKI cases separately from diabetes complication cases. We calculated PPVs (true positives/flagged cases) and associated 95% CIs. Additionally, we conducted detailed examination of false positives to understand why they were flagged, as well as to determine how to improve the PSI. We determined whether false positives resulted from cases being coded inappropriately or from coding limitations (ie, despite correct coding, the case did not meet the clinical intent of the indicator) (Table 1).

We compared true positives and false positives with respect to selected demographics and assigned codes, using *t*-tests for continuous variables and chi-squares for categorical variables. We also examined true positives to determine the clinical consequences of PMD and factors contributing to its occurrence, and we performed descriptive analyses of relevant variables. Analyses were performed using SAS software, version 9.1 (SAS Institute Inc).

Table 1. Characteristics of Sample Patients

Variable	All AKI patients n = 104	TP AKI n = 73	FP AKI n = 31	All DM patients n = 15	TP DM n = 2	FP DM n = 13
Age, y, mean (SD)	67.8 (10.5)	69.0 (9.7)	65.2 (12.1)	56.4 (12.2)	55.0 (1.4)	56.6 (13.2)
Gender, male, n (%)	102 (98.1)	72 (98.6)	30 (96.8)	15 (100)	2 (100)	13 (100)
Race, n (%)						
White	74 (71.2)	53 (72.6)	21 (67.7)	10 (67.7)	1 (50.0)	9 (69.2)
African-American	13 (12.5)	10 (13.7)	3 (9.7)	3 (20.0)	–	3 (23.1)
Hispanic	8 (7.7)	4 (5.5)	4 (12.9)	1 (6.7)	–	1 (7.7)
Other/missing	9 (8.7)	6 (8.2)	3 (9.7)	1 (6.7)	1 (50.0)	–
ICD-9-CM codes - AKI diagnosis codes, n (%)*						
584.x only	76 (73.1)	53 (72.6)	23 (74.2)	–	–	–
Both 584 and 997.5	13 (12.5)	10 (13.7)	3 (9.7)	–	–	–
997.5 only	7 (6.7)	6 (8.2)	1 (3.2)	–	–	–
586 only	8 (7.7)	4 (5.5)	4 (12.9)	–	–	–

* ICD-9-CM codes: 584.x,"acute renal failure"; 997.5, "urinary complications" includes acute renal failure due to a procedure; 586,"renal failure unspecified." All diabetes patients had diabetic ketoacidosis codes (250.10 to 250.13).
Percentages represent column percents. Cases were flagged per the patient safety indicator algorithm.
There were no significant differences between TPs and FPs among AKI or diabetes patients. Diabetes patients were significantly younger than AKI patients ($p < 0.05$).
AKI, acute kidney injury; DM, diabetes mellitus; FP, false positive; TP, true positive.

RESULTS

PPV for PMD indicator

Of 86,321 eligible hospitalizations among sample hospitals, 185 were flagged for PMD (an observed rate of 2.1 per 1,000), which was comparable to the national VA flagged PMD rate (578 flagged cases per 268,552 eligible hospitalizations, or 2.1 per 1,000). Of the 119 reviewed cases, 75 were true positives, for a PPV of 63% (95% CI 54% to 72%). Of 104 AKI cases, 73 were true positives, for a PPV of 70% (95% CI 60% to 79%); of the 15 diabetes cases, only 2 were true positives, for a PPV of 13% (95% CI 2% to 40%).

Table 1 shows demographic and coding characteristics of all flagged cases, true positives, and false positives. Both true positive and false positive patients with AKI were older than patients flagged with diabetes complications (mean ± SD, 67.8 ± 10.5 years vs. 56.4 ± 12.2 years; $p < 0.001$). Of the AKI sample, 73% of cases were flagged based on a 584.x diagnosis code ("acute renal failure"), 8% based on a 586 code ("renal failure unspecified"), 7% based on a 997.5 code ("urinary complications," includes acute renal failure due to a procedure); the remainder had a combination of a 584x and a 997.5 code. All of the flagged diabetes cases had a code for diabetic ketoacidosis. See Appendix 1 for further information on code definitions.

False positive PMD cases

AKI cases

Of the 31 false positive AKI cases, 4 (13%) failed to meet coding criteria; they were all patients with end-stage renal disease (ESRD) present on admission, who received a 584.x code for unclear reasons. An additional 27 cases were coded appropriately, but failed to meet the indicator's clinical intent. Twenty-one cases were nonelective admissions (68%); 4 cases (13%) had ESRD present on admission; however, these were assigned less specific codes (586 or 997.5), so we did not consider them as coding errors. An additional 2 elective cases (6%) did have postoperative AKI requiring dialysis. However, 1 was an anticipated outcome of a procedure: a patient with a solitary kidney who underwent nephrectomy for a renal mass. In the other, the procedure was incidental to the development of AKI: a patient postcardiac transplantation with hepatitis C admitted for investigation of a cavitary lung nodule and elevated transaminases. He underwent a palate biopsy for a mouth ulcer and subsequently developed renal failure thought to be due to hepatorenal syndrome and/or from amphotericin B, which he was receiving for pulmonary aspergillosis.

Two nonelective cases would also have been excluded based on AKI being present on admission and having no preceding procedure. The 4 cases that failed to meet coding criteria would also have been excluded by other criteria because they all had ESRD present on admission (3 of these patients were actually admitted for dialysis access procedures).

Diabetes complications cases

Among the 13 false positive diabetes cases, there were no coding errors; all failed to meet the clinical intent of the

indicator. Ten were nonelective (69%); 9 of these were also present on admission (and unrelated to previous surgery). Two patients had ESRD present on admission (this is an exclusion criterion for both diabetes and AKI patients), and 1 patient had diabetic ketoacidosis incorrectly documented in the discharge summary as a secondary diagnosis on the current admission when it had occurred during a previous admission (and was unrelated to a procedure).

True positive PMD cases

AKI cases

Because only 2 of 75 true positives had diabetes complications, we present summary results only for the AKI true positives. See Appendix 2 online for details on the 2 true positive diabetes cases.

Outcomes

Median length of stay was 30 days (25th, 75th percentile: 15, 51 days, respectively). Nine patients underwent continuous hemofiltration, the remainder underwent standard intermittent hemodialysis. Nineteen true positive patients (26%) were discharged on dialysis. There were 27 in-hospital deaths (37%). The median time between the initial operation and initiation of dialysis was 7 days (25th, 75th percentile: 3, 15 days, respectively).

Procedure-related risk factors (Table 2)

Of the true positives, 30% of cases (n = 22) were associated with cardiac index procedures. This included 16 cardiac and 3 thoracic aorta procedures, all performed on cardiopulmonary bypass, and 3 coronary angioplasties; median on-pump time was 181 minutes, range 144 to 375 minutes, in the 10 cases in which this was documented. Twenty-one percent (n = 15) of associated index procedures were vascular, involving the abdominal aorta or lower extremities (12 of these involved abdominal aorta clamping; 4 were clamped suprarenally). Sixty-four true positive patients (88%) developed AKI after the first operating room (index) procedure; in 6 (8%) this followed a second procedure, and in 3 (4%), a third procedure. For all but 1 of the 9 true positive patients with AKI after a subsequent procedure, this was a nonelective procedure. Most (93%) index procedures were performed under general or epidural anesthesia. Thirty-seven noncardiac patients (51% of all patients and 73% of noncardiac patients) underwent an operation lasting longer than 3 hours. (All on-pump cardiac cases were longer than 3 hours.) Twenty noncardiac patients (27%) lost at least 1L of blood intraoperatively. Forty-one percent (n = 30) received at least 3 L of intravenous fluid intraoperatively (crystalloid, colloid; although this information was available on only 42 patients [58%]), and 48% (n = 35) received the equivalent of at least 1 unit of blood). (See Table 2 for further details.)

Table 2. Procedure-Related Risk Factors

Variable	True positives (n = 73)
Anatomic region of index procedure, n (%)	
Cardiac	22 (30.1)
On-pump procedures*	19 (26.0)
PTCAs	3 (4.1)
Vascular†	15 (20.5)
Abdominal/gastrointestinal	17 (23.3)
Urologic‡	9 (12.3)
Orthopaedic	5 (6.8)
Respiratory	5 (6.8)
Type of anesthesia, n (%)	
General only	51 (69.9)
General + epidural	17 (23.3)
Epidural only	2 (2.7)
Unknown§	3 (4.1)
Operating room blood loss, mL, median (range)	800 (10–8,700)
Operating room fluid input, mL, median (range)	4,600 (0–14,000)
Duration of operation, min, median, (range)	317 (76–695)
Number of operating room procedures before AKI,‖ n (%)	
1	64 (87.7)
2	6 (8.2)
3	3 (4.1)

*Includes 3 thoracic aorta repairs (2 ascending repairs with aortic valve replacement, 1 descending – all done on pump); the remainder were coronary artery bypass grafts and/or valve repairs.
†Includes 6 open abdominal aortic aneurysm repairs, 6 bypass procedures involving the abdominal aorta; 1 endovascular abdominal aortic aneurysm repair and 2 lower extremity bypass procedures.
‡Includes 7 nephrectomies, including one patient with a single kidney.
§These were 3 coronary angioplasties.
‖In all but 3 of the 9 cases with acute kidney injury after a subsequent procedure, these procedures resulted from complications directly associated with the initial operation. The 3 cases were: an open cholecystectomy after a coronary artery bypass grafts /mitral valve replacement, a small bowel resection after a coronary angioplasty, and a lower extremity fasciotomy after esophagectomy.
PTCA, percutaneous transluminal coronary angioplasty.

Patient-related risk factors

Thirty-seven percent of patients had a history of diabetes present on admission; 23% had congestive heart failure. Although 33% had a diagnosis of chronic kidney disease documented per the preoperative anesthesia or admission note, 55% had an estimated glomerular filtration rate less than 60 mL/minute on admission (based on the modified Modification of Diet in Renal Disease [MDRD] formula); it was less than 30 mL/minute in 10% (Table 3).

Table 3. Patient-Related and Perioperative Risk Factors

Variable	True positives (n = 73)
BMI, mean (SD), kg/m^2	29.9 (7.5)
Conditions present on admission, n (%)	
Diabetes	27 (37.0)
Congestive heart failure	17 (23.3)
Documented chronic kidney disease	24 (32.9)
Admission creatinine, mg/dL, mean (SD)	1.5 (0.6)
Admission GFR, mg/dL, mean (SD)*	59.4 (19.7)
Admission GFR < 60 mL/min, n (%)	40 (54.8)
Drugs potentially affecting renal function†	
Aminoglycosides, vancomycin	6 (8.2)
ACEIs/ARBs	18 (24.7)
NSAIDs	2 (2.7)
Contrast‡	18 (24.7)

*Based on the modified Modification of Diet in Renal Disease (MDRD) equation.
†Received during admission, before diagnosis of acute kidney injury.
‡This included 3 coronary angiograms with angioplasties, 2 intraoperative arteriograms (both aortobifemoral bypasses with femoral embolectomies), 2 preoperative aorta and lower extremity angiograms, and 3 MRIs with gadolidium; the remainder were CT scans, including 2 CT angiograms. The contrast type was nonionic in 8 cases and unknown in 8 cases.
ACEIs/ARBs, angiotensin converting enzyme inhibitors/angiotensin receptor blockers; BMI, body mass index; GFR, glomerular filtration rate; NSAIDs, nonsteroidal anti-inflammatory drugs.

Contributing factors

The most common nephrologist-identified contributing causes were related to renal hypoperfusion (84%), most often due to perioperative hypotension (45%) or sepsis (34%). Although renal ischemia was often noted as a cause, it was only explicitly mentioned as due to renal artery or aorta occlusion in 4 vascular surgery cases. Exogenous nephrotoxins were implicated in 22% of cases; this was split between contrast (n = 8) and other drugs such as antibiotics (n = 8). Only 1 of 6 patients on an aminoglycoside antibiotic or vancomycin had this noted as a possible contributing factor. Virtually all patients had more than 1 presumed cause contributing to AKI (Table 4).

In terms of preventive measures, 8 of 18 patients who received intravenous or intra-arterial contrast before AKI onset had an estimated glomerular filtration rate less than 60 mL/minute; only 3 of these 8 patients (38%) had documentation of receipt of prophylaxis such as hydration with normal saline or sodium bicarbonate or use of n-acetylcysteine. (Two nonprophylaxed chronic kidney disease patients developed contrast-induced AKI, as did 2 prophylaxed chronic kidney disease patients who also received nonionic isomolar agents.)

DISCUSSION

This is one of the first studies to examine the criterion validity of the AHRQ PSI, PMD in its current form in any setting.[17] Although the overall indicator's PPV was 63%, we found that diabetes complications were rare and were associated with a very low PPV (13%). Conversely, AKI-related complications had a moderate PPV (70%). This latter value is comparable to the upper reported PPV range of other postoperative PSIs validated in non-VA settings.[27,28] Our PPV for the AKI-subset findings was slightly higher than that, supporting the previously reported value of 63% found in the VA using 2001 data.[17] Our findings also support the addition of the less specific renal failure codes (586 and 997.5) to the current version of the indicator.[17] Had we excluded them, our PPV would have been essentially unchanged at 71%, but we would have missed 10 true positives. To our knowledge, no other data exist with respect to the validity of coding for postoperative diabetes complications included in this PSI. There are also no data specific to diabetic ketoacidosis or hyperosmolar coding in any inpatient setting. However, a previous study found low coding rates and poor PPV for hypoglycemic diabetes complications.[29] Geraci and colleagues[29] were able to confirm only 35% (12 of 34) of coded hypoglycemic episodes in medical patients by chart review; conversely only 12 of 76 (16%) chart-identified hypoglycemic epi-

Table 4. Nephrologist-Presumed Causes of Postoperative Acute Kidney Injury

Factor*	n	%
Renal hypoperfusion	61	83.6
Decreased effective intravascular volume (congestive heart failure, hepatorenal syndrome)	14	19.2
Intravascular volume depletion (blood loss, diuretics)	12	16.4
Sepsis	25	34.7
Hypotension/cardiac arrest	33	45.2
Renal artery occlusion/clamping of aorta†	4	5.5
Abdominal compartment syndrome	2	2.7
Nephrotoxins	24	33.3
Exogenous nephrotoxins	16	21.9
Contrast	8	11.0
Other nephrotoxins‡	8	11.0
Endogenous nephrotoxins (myoglobin)§	10	13.7
Cholesterol emboli/atheroemboli	11	15.1
Other‖	7	9.6

*Multiple factors may be responsible for a given case. Causes were abstracted from nephrology notes. There is also overlap between some of the renal hypoperfusion causes (eg, hypotension may be associated with any of the listed causes).
†12 true positive patients had cross-clamping of the abdominal aorta; only 4 had it specifically noted as a contributing cause of acute kidney injury.
‡This included 6 cases of antibiotic presumed acute interstitial nephritis.
§These were all due to myoglobinuria, and most frequently associated with lower extremity reperfusion after revascularization procedures. In 2 cases, a statin and a fibrate were thought to contribute. These 2 cases were also counted under "Other nephrotoxins."
‖Seven cases specify the patient had a picture consistent with acute tubular necrosis but no additional details are given.

sodes were coded as such. This is noteworthy, given that CMS has adopted a related measure, "manifestations of poor glycemic control," which is applicable to all inpatient discharges, as one of its hospital-acquired conditions.[30]

Suggested coding improvements to the PMD indicator

With respect to false positives, coding errors were relatively rare, with most false positives representing cases that did not satisfy the PSI's clinical definition (ie, patients with nonelective hospitalizations or present-on-admission conditions such as ESRD and diabetic ketoacidosis). As mentioned, emergency admissions, a proxy for emergency procedures, are intended to be excluded. Because the VA lacks an admission status code, we used an algorithm designed to distinguish nonelective admissions, which incorrectly identified 21 flagged AKI cases and 10 diabetes-related cases as elective (26% overall).[19] However, even admission status is not a perfect proxy for procedure status; 3 AKI patients with nonelective admissions actually underwent elective procedures. Further, it is possible that some PMD cases associated with nonelective cases might be preventable. If we kept cases classified as false positives based solely on admission status, the PPV for diabetes-related complications would increase only to 20%. However, the AKI-related PPV would increase to 88% (95% CI 76% to 91%). Of note, even in settings that code data for admission status, such codes may not be always be applied properly. A non-VA study of the PSI "postoperative respiratory failure" (which similarly is intended to flag cases associated with elective admissions) found incorrect coding of elective status to be the most frequent reason for false positives.[28] So future research should examine the effect of expanding PSI denominator exclusions to additional principal diagnoses that are "suspicious for a nonelective presentation" (eg, sepsis).[28]

Although the VA does not code for present-on-admission status, in settings that have such coding, reported present-on-admission rates for PMD (AKI and diabetes complications combined) have ranged from 9% using chart validation at a single hospital[31] to 36% based on New York State data.[32] If the VA used present-on-admission codes, and these were applied correctly, 73% of diabetes cases but only 10% of AKI cases would have been excluded.

Based on our false positive review, we suggest some relatively straightforward modifications to improve this indicator's PPV. First, due to the low rate, poor PPV, and existing data showing infrequent coding of diabetes-related complications,[29] in addition to pathophysiologic differences, we believe the indicator should be restricted to AKI cases. Given CMS' related hospital-acquired condition measure, "manifestations of poor glycemic control," further efforts should be devoted to improving coding of diabetes complications. The PPV of the AKI cases in the VA could be improved by adoption of admission status and present-on-admission codes. Appropriate use of such codes would have eliminated all of our AKI false positives. VA adoption of an admission status code would also allow more meaningful VA and non-VA comparisons of PMD rates. Further, 3 false positives were patients with ESRD admitted for a principal procedure related to dialysis access (39.27, "arteriovenostomy for renal dialysis;" 39.41, "revision of arteriovenous shunt for dialysis;" and 39.93, "insertion of vessel-to-vessel cannula"). Such principal procedures could be excluded from the denominator.

We also had 1false positive patient who had his remaining kidney removed because of a renal mass. Although this patient was appropriately coded, postoperative AKI was an anticipated outcome of care, as opposed to a complication. However, there currently is no way to exclude such cases based on coding. One approach would be to add the term *anephric* to existing ICD-9-CM diagnosis codes, and exclude cases with this diagnosis and a principal procedure of a nephrectomy.

PMD preventability

PMD is intended to detect complications that are avoidable through good care. Although studies of other postoperative indicators provide empirical evidence of this, there are no data specific to PMD.[33] So, as a secondary study goal, we explored the preventability of true events through abstracted information on risk factors, processes of care, and causes. (Given we are recommending removal of diabetes cases from the indicator, this discussion is restricted to AKI prevention.) As expected, most of our AKI cases were due to acute tubular necrosis, representing the end result of multiple factors related to renal hypoperfusion or renal damage from nephrotoxins.[34] Because there is little that can be done to reverse postoperative AKI once a renal insult occurs,[35] management is primarily preventive (limiting perioperative exposure to renal hypoperfusion and nephrotoxins). However, there are few specific evidence-based methods for prevention, except perhaps with respect to contrast-induced nephropathy.[36,37] Preoperative evaluation is recommended to recognize and if possible modify predisposing factors; prediction models can help identify particularly high risk patients who may require more intensive perioperative monitoring of parameters including vital signs, fluid status, and cardiac hemodynamics.[3,4,9,34] These latter methods seem clinically intuitive, but there are no trials to support their use. Other measures, such as preoperative evaluation and modification of cardiac and pulmonary risks, and appropriate infection prophylaxis, may indirectly reduce AKI risk.

Although many of our patients appeared to be at high risk based on known procedure- (cardiac and abdominal aortic procedures) and patient-related risk factors (advanced age, chronic kidney disease, congestive heart failure, diabetes),[9,34] the extent to which these risk factors could have been modified preoperatively is unclear. Most patients had documentation of a preoperative assessment by anesthesia at a minimum and many were followed by cardiology, but it was beyond this study's scope to ascertain whether preoperative assessments resulted in changes in the management of comorbidities or choice of procedure that might have decreased risk.

In terms of processes of care related to nephrotoxin exposure, only 38% of patients with chronic kidney disease who received contrast were given any sort of prophylaxis including intravenous hydration. With respect to perioperative processes related to limiting hypoperfusion, although longer cardiopulmonary bypass on-pump and aortic cross-clamping times increase AKI risk,[2] it was not clear whether anything could have been done to shorten these times. We also did not have access to intraoperative anesthesia records, so we do not know whether patients experienced intraoperative blood pressure fluctuations, or the degree to which such information would help to assess preventability. We do know that all cardiac and vascular patients, as well as several other high-risk patients, were closely monitored perioperatively in the ICU. Finally, we did not abstract information on processes of care less directly related to AKI, such as infection prevention; we addressed this issue in detail in a study of the PSI "postoperative sepsis."[38]

Study limitations and strengths

We lack information on the indicator's sensitivity or specificity because abstracting cases that were not flagged by the PSI algorithm was beyond the scope and resources of this study. Additionally, this was a sample of mainly elderly men. In spite of this, our findings with respect to risk factors and outcomes were consistent with those in existing literature.[9,34] Though a few patients received contrast without appropriate prophylaxis, we cannot determine whether any other cases were associated with potential quality of care problems. This is partly because we used retrospective chart-based data, which rely on completeness of documentation, the relatively small sample size, the absence of a control group, and lack of certain data that are unavailable in VistAWeb. Further, the abstraction tool was designed to be as explicit as possible; the nurses were advised against drawing inferences about causation to maximize the reliability of findings.

In terms of study strengths, we randomly selected cases from a nationwide sample of VA hospitals. We performed IRR testing of data abstraction both early and late in the process, and attained high abstractor agreement. We also conducted a thorough examination of false positives and were able to identify ways to improve the PPV of the indicator. Also of note, PMD is part of the CMS-tracked PSI composite measure and overlaps with the diabetes-related hospital-acquired condition, both of which are being linked to sanctions.[12,30] Although this issue could potentially influence the accuracy of both coding and medical record documentation in Medicare-reimbursed settings, it is unlikely to have affected our findings because neither measure is currently tracked in the VA.

CONCLUSIONS

In its present state, PMD should continue to be used as a screen for patient safety events as opposed to a performance measure. However, to improve the usefulness of this indicator for detecting true events, we strongly recommend removing the diabetes-related complications from the indicator and focusing on AKI-related complications. The VA should also adopt admission status and present-on-admission codes. Further research regarding the extent to which PMD identifies events that are preventable through improved care is also necessary before it can be considered a definitive quality measure. Results of similar PMD validation efforts being conducted by AHRQ investigators in the non-VA setting should soon be available for comparison.

Author Contributions

Study conception and design: Borzecki, Shin, Itani, Rosen
Acquisition of data: Borzecki, Chen
Analysis and interpretation of data: Borzecki, Cevasco, Chen, Itani, Rosen
Drafting of manuscript: Borzecki
Critical revision: Borzecki, Cevasco, Chen, Shin, Itani, Rosen

Acknowledgment: We acknowledge Haytham Kaafarani, MD, MPH, for assistance with abstraction instrument and guidelines modification; Kathleen Hickson, RN, MA, and Sally MacDonald, RN for their assistance with abstraction instrument and guidelines modification, data abstraction, and data cleaning; Hillary Mull, MPP, for her technical assistance in developing the electronic abstraction form; Dijon Fasoli, RN, PhD for conducting inter-rater reliability assessments; and Susan Loveland, MAT for programming expertise and assistance with identifying the hospital sample and flagged cases.

REFERENCES

1. Agency for Healthcare Research and Quality. AHRQ Quality Indicators: Guide to Patient Safety Indicators. Rockville, MD; March 2003; Version 3.1 (March 12, 2007)
2. Sear JW. Kidney dysfunction in the postoperative period. Br J Anaesth 2005;95:20–32.

3. Thakar CV, Arrigain S, Worley S, et al. A clinical score to predict acute renal failure after cardiac surgery. J Am Soc Nephrol 2005; 16:162–168.
4. Chertow GM, Lazarus JM, Christiansen CL, et al. Preoperative renal risk stratification. Circulation 1997;95:878–884.
5. Svensson LG, Coselli JS, Safi HJ, et al. Appraisal of adjuncts to prevent acute renal failure after surgery on the thoracic or thoracoabdominal aorta. J Vasc Surg 1989;10:230–239.
6. Mangano CM, Diamondstone LS, Ramsay JG, et al. Renal dysfunction after myocardial revascularization: risk factors, adverse outcomes, and hospital resource utilization. The Multicenter Study of Perioperative Ischemia Research Group. Ann Intern Med 1998;128:194–203.
7. Dimick JB, Pronovost PJ, Cowan JA, Lipsett PA. Complications and costs after high-risk surgery: where should we focus quality improvement initiatives? J Am Coll Surg 2003;196:671–678.
8. McHugh MD, Shang J, Sloane DM, Aiken LH. Risk factors for hospital-acquired 'poor glycemic control': a case-control study. Int J Qual Health Care 2010;23:44–51.
9. Borthwick E, Ferguson A. Perioperative acute kidney injury: risk factors, recognition, management, and outcomes. BMJ 2010; 341:c3365.
10. Standards of medical care in diabetes–2009. Diabetes Care 2009;32 (Suppl 1):S13–61.
11. National Quality Forum. NQF endorsed standards. Available at: http://www.qualityforum.org/Measures_List.aspx. Accessed June 1, 2010.
12. RHQDAPU Program Overview: Reporting hospital quality data for annual payment update. Available at: http://www.qualitynet.org/dcs/ContentServer?cid=1138115987129&pagename=QnetPublic%2FPage%2FQnetTier2&c=Page. Accessed March 1, 2010.
13. CMS. Fiscal year 2009 quality measure reporting for 2010 payment update. Available at: http://www.cms.hhs.gov/HospitalQualityInits/downloads/HospitalRHQDAPU200808.pdf. Accessed August 1, 2010.
14. McDonald KM, Romano PS, Geppert J, et al. Measures of patient safety based on hospital administrative data - the patient safety indicators. Technical review 5 (Prepared by the UCSF-Stanford Evidence-Based Practice Center under Contract # 290-97-0013). AHRQ Pub. No. 02-0038. Rockville, MD: Agency for Health Care Research and Quality; August 2002.
15. Rosen AK, Rivard P, Zhao S, et al. Evaluating the patient safety indicators: how well do they perform on Veterans Health Administration data? Med Care 2005;43:873–884.
16. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA 2003;290:1868–1874.
17. Romano PS, Mull HJ, Rivard PE, et al. Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. Health Serv Res 2009;44:182–204.
18. VA Information Resource Center. VIReC Research User Guide: FY2002 VHA Medical SAS Inpatient Datasets. Available at: http://www.virec.research.va.gov/References/RUG/RUG-Inpatient02.pdf. Accessed December 1, 2009.
19. Rivard P, Elwy AR, Loveland S, et al. Applying patient safety indicators (PSIs) across healthcare systems: Achieving data comparability. Vol 2:7-25. Rockville, MD: Agency for Healthcare Research and Quality. AHRQ Publication No. 005-0021-CD; 2005.
20. Veterans Health Administration. VistAWeb Overview. Available at: http://vista.med.va.gov/vistaweb/. Accessed January 10, 2010.
21. Agency for Healthcare Research and Quality. Patient safety indicators: Technical specifications. March 2003; Version 3.1 (March 12, 2007).
22. Borzecki AM, Kaafarani HMA, Utter GH, et al. How valid is the AHRQ patient safety indicator "postoperative respiratory failure"? J Am Coll Surg 2011;212:935–945.
23. AHRQ. Patient safety indicators:software documentation, version 3.2 SAS March 2003 Version 3.2 (March 10, 2008).
24. Agency for Healthcare Research and Quality. AHRQ quality indicators: patient safety indicator composite SAS documentation version 3.1a (March, 2007).
25. Lawthers AG, McCarthy EP, Davis RB, et al. Identification of in-hospital complications from claims data. Is it valid? Med Care 2000;38:785–795.
26. AHRQ quality indicators validation pilot. Available at: http://qualityindicators.ahrq.gov/validationpilot.htm. Accessed March 1, 2010.
27. White RH, Sadeghi B, Tancredi DJ, et al. How valid is the ICD-9-CM based AHRQ patient safety indicator for postoperative venous thromboembolism? Med Care 2009;47:1237–1243.
28. Utter GH, Cuny J, Sama P, et al. Detection of postoperative respiratory failure: how predictive is the Agency for Healthcare Research and Quality's patient safety indicator? J Am Coll Surg; 211:347–354 e341–329.
29. Geraci JM, Ashton CM, Kuykendall DH, et al. International Classification of Diseases, 9th Revision, Clinical Modification codes in discharge abstracts are poor measures of complication occurrence in medical inpatients. Med Care 1997;35:589–602.
30. CMS hospital-acquired condition (HAC) measures for RHQDAPU. Measure specifications. Available at: www.ihaconnect.org/about/moreinformation/unlockeddocs/d66248.aspx. Accessed December 1, 2010.
31. Bahl V, Thompson MA, Kau TY, et al. Do the AHRQ patient safety indicators flag conditions that are present at the time of hospital admission? Med Care 2008;46:516–522.
32. Houchens RL, Elixhauser A, Romano PS. How often are potential patient safety events present on admission? Jt Comm J Qual Patient Saf 2008;34:154–163.
33. Weingart SN, Iezzoni LI, Davis RB, et al. Use of administrative data to find substandard care: validation of the complications screening program. Med Care 2000;38:796–806.
34. Carmichael P, Carmichael AR. Acute renal failure in the surgical setting. ANZ J Surg 2003;73:144–153.
35. Zacharias M, Gilmore IC, Herbison GP, et al. Interventions for protecting renal function in the perioperative period. Cochrane database of systematic reviews (Online) 2005(3): CD003590.
36. Kelly AM, Dwamena B, Cronin P, et al. Meta-analysis: effectiveness of drugs for preventing contrast-induced nephropathy. Ann Intern Med 2008;148:284–294.
37. Zoungas S, Ninomiya T, Huxley R, et al. Systematic review: sodium bicarbonate treatment regimens for the prevention of contrast-induced nephropathy. Ann Intern Med 2009;151: 631–638.
38. Borzecki AM. Validating the patient safety indicators in the Veterans Health Administration. AcademyHealth Annual Meeting 2009, Chicago, IL. Available at: http://www.academyhealth.org/files/2009/tuesday/borzecki.pdf. Accessed October 10, 2010.

Appendix 1. Postoperative Physiologic and Metabolic Derangement Indicator: Detailed Definition & ICD-9-CM Codes

Definition	Cases of specified physiologic or metabolic derangement procedure per 1,000 elective surgical discharges with an operating room (OR) procedure*
Numerator	Discharges among cases meeting the inclusion and exclusion rules for the denominator with ICD-9-CM codes for physiologic and metabolic derangements in any secondary diagnosis field. Includes the following: ● Acute renal failure (584.5–584.9 – acute renal failure, 586 – renal failure not otherwise specified, 997.5 – urinary complications, including acute renal failure complicating a procedure) accompanied by a procedure code for dialysis (39.95 - hemodialysis, 54.98 – peritoneal dialysis) *or* ● Diabetes (type 1 or 2) with: ○ ketoacidosis (250.10–250.13), or ○ hyperosmolarity (250.20–250.23), or ○ coma (250.30-250.33)
Denominator	All elective surgical discharges age 18 and over defined by specific DRGs and an ICD-9-CM code for an OR procedure.† Exclude cases with: ● A preexisting (principal diagnosis or secondary diagnosis present on admission, if known) of physiologic and metabolic derangements OR chronic renal failure‡ ● Acute renal failure where a procedure for dialysis occurs before or on the same day as the first OR procedure ● Both a diagnosis code of ketoacidosis, hyperosmolarity, or other coma (subgroups of physiologic and metabolic derangements coding) AND a principal diagnosis of diabetes ● Both a secondary diagnosis code for acute renal failure (subgroup of physiologic and metabolic derangements coding) and a principal diagnosis of acute myocardial infarction, cardiac arrhythmia, cardiac arrest, shock, hemorrhage, or gastrointestinal hemorrhage ● MDC 14 (pregnancy, childbirth and the puerperium)

*Reported as rates, risk-adjusted for age, sex and modified Diagnosis Related Groups (DRGs).

†Elective procedure defined by admit type. See PSI technical specifications for the complete list of eligible surgical DRG and operating room (OR) procedure codes. (Agency for Healthcare Research and Quality. Patient Safety Indicators: Technical Specifications. March 2003; Version 3.1 [March 12, 2007].) Eligible OR procedures include all procedures that are considered valid OR procedures by the DRG grouper; procedures such as a diagnostic cardiac catheterization are excluded, while a percutaneous coronary angioplasty is included.

‡The actual PSI SAS code specifies that this is advanced chronic renal failure (stage V) or end-stage renal disease.

MDC = Major Diagnostic Category; this is based on the principal diagnosis.

Appendix 2. True Positive Diabetes Complications Cases

Both cases were obese patients on short and long-acting insulin combined with oral agents prior to admission, admitted for major surgery, with endocrinologist-diagnosed probable diabetic ketoacidosis (DKA) noted on postoperative day (POD) 3. The first case did not receive perioperative basal insulin, the second case did.

Case 1 – This was a 54 year-old obese male, with a 30-year history of diabetes, admitted for a total knee replacement. He was on a sulfonylurea, glitazone, rapid-acting (glulisine) and long-acting insulin (glargine – 50U daily) prior to admission with reasonably-controlled sugars (A1c 7.6% on POD 3).

He was admitted the same day as the procedure. He took his usual medications the day before. He received IV ½ NS pre- and post-operatively without D5W. He did not receive any basal insulin or oral medications post-operatively till he developed DKA; he was covered only with regular insulin per sliding scale, starting with a glucose ≤200 mg/dl. He started to develop progressive elevation in his sugars late on POD 1 coinciding with advancing his diet. On POD 1 his sugars were up to 180 mg/dl, by POD 2 they were >300 mg/dl. On POD 3, he developed nausea and vomiting in the early am; serum glucose was 402 mg/dl, CO_2 was 9 meq/L, and urine was positive for ketones. Endocrinology was consulted; the patient was diagnosed with DKA, and transferred to the intensive care unit for IV insulin. (It took 4 days to clear the ketones.) No reason for the DKA was noted. Presumably the patient developed ketoacidosis because of insulin deficiency and this might have been avoided were he covered with basal insulin.

Case 2 – This was a 56 year-old obese male, with an 8-year history of diabetes, admitted for a right hemicolectomy. He was on a glitazone, intermediate (NPH, 40U am and 50U pm) and short-acting insulin (twice daily) prior to admission, with poorly controlled sugars (>300; A1c 12.0% on the day of admission). Endocrinology was involved in his perioperative care from POD 1 onward.

He was admitted the day before the surgery and was given IV ½ NS with D5W perioperatively. In hospital he received maintenance insulin, initially with NPH subcutaneously (sc), then with an insulin drip, and then with NPH twice daily supplemented with short-acting insulin sc with meals up until possible DKA was diagnosed in the setting of an acute gouty flare; his blood sugar at the time was 267 mg/dl. However, it is not clear he actually had DKA because he only had one urinalysis with a small amount of ketones and had a CO_2 of 28 meq/L. Per endocrinology's

recommendations, he was managed only with an increase in his NPH and regular insulin twice daily, and not with IV insulin. However, endocrinology never wrote a subsequent note to say he didn't have DKA. (The ketonuria may have been due to starvation, since he had just started a clear diet that day after a couple of days of fasting.) This case was coded appropriately, but one could argue the clinical evidence did not support the diagnosis of DKA.

How Valid is the AHRQ Patient Safety Indicator "Postoperative Hemorrhage or Hematoma"?

Ann M Borzecki, MD, MPH, Haytham Kaafarani, MD, MPH, Marisa Cevasco, MD, MPH, Kathleen Hickson, RN, MA, Sally MacDonald, RN, Marlena Shin, JD, MPH, Kamal MF Itani, MD, FACS, Amy K Rosen, PhD

BACKGROUND: Postoperative hemorrhage or hematoma (PHH), an Agency for Healthcare Research and Quality Patient Safety Indicator, uses administrative data to detect cases of potentially preventable postsurgical bleeding requiring a reparative procedure. How accurately it identifies true events is unknown. We therefore determined PHH's positive predictive value.

STUDY DESIGN: Using Patient Safety Indicator software (v.3.1a) and fiscal year 2003–2007 discharge data from 28 Veterans Health Administration hospitals, we identified 112 possible cases of PHH. Based on medical record abstraction, we characterized cases as true (TPs) or false positives (FPs), calculated positive predictive value, and analyzed FPs to ascertain reasons for incorrect identification and TPs to determine PHH-associated clinical consequences and risk factors.

RESULTS: Eighty-four cases were TPs (positive predictive value, 75%; 95% CI, 66−83%); 63% had a hematoma diagnosis, 30% had a hemorrhage diagnosis, 7% had both. Reasons for FPs included events present on admission (29%); hemorrhage/hematoma identified and controlled during the original procedure rather than postoperatively (21%); or postoperative hemorrhage/hematoma that did not require a procedure (18%). Most TPs (82%) returned to the operating room for hemorrhage/hematoma management; 64% required blood products and 7% died in-hospital. The most common index procedures resulting in postoperative hemorrhage/hematoma were vascular (38%); 56% were performed by a physician-in-training (under supervision). We found no substantial association between physician training status or perioperative anticoagulant use and bleeding risk.

CONCLUSIONS: PHH's accuracy could be improved by coding enhancements, such as adopting present on admission codes or associating a timing factor with codes dealing with bleeding control. The ability of PHH to identify events representing quality of care problems requires additional evaluation. (J Am Coll Surg 2011;212:946–953. © 2011 by the American College of Surgeons)

The Agency for Healthcare Research and Quality (AHRQ) Patient Safety Indicators (PSIs) were developed in response to demand for easily applied measures that could guide quality-improvement initiatives and monitor trends in patient safety.[1] Because they use hospital administrative discharge abstracts, they were originally intended as screens for potentially preventable inpatient complications, highlighting areas where quality of care should be investigated rather than being definitive measures. However, the National Quality Forum recently endorsed several PSIs as hospital performance measures, and the Centers for Medicare and Medicaid Services (CMS) are adding 4 individual PSIs and a composite PSI to measures tracked by their hospital reporting initiative.[2-4] Underperforming or nonreporting hospitals will receive reduced payments.

The increasing use of PSIs as a measure of quality and safety requires users to understand their strengths and lim-

Disclosure Information: Nothing to disclose.

The study was funded by a Department of Veterans Affairs Health Services Research and Development service grant number SDR 07-002.

Abstract presented at the annual AcademyHealth meeting, Boston, MA, June 2010.

Received August 18, 2010; Revised September 16, 2010; Accepted September 17, 2010.

From the Center for Health Quality, Outcomes and Economic Research, Bedford VAMC, Bedford (Borzecki, Hickson, MacDonald), Department of Health Policy and Management, Boston University School of Public Health (Borzecki, Rosen), Boston University School of Medicine (Borzecki, Itani), Department of Surgery, Tufts Medical Center, Tufts University School of Medicine (Kaafarani), Department of Surgery, VA Boston Healthcare System (Cevasco, Itani), Department of Surgery, Brigham and Women's Hospital (Cevasco), Center for Organization, Leadership, and Management Research, VA Boston Healthcare System (Shin, Rosen), Harvard Medical School (Itani), Boston, MA.

Correspondence address: Ann M Borzecki, MD, MPH, CHQOER, Bedford VA Hospital (152), 200 Springs Rd, Bedford, MA 01730. email: amb@bu.edu

ISSN 1072-7515/11/$36.00
doi:10.1016/j.jamcollsurg.2010.09.033

Abbreviations and Acronyms

AHRQ = Agency for Healthcare Research and Quality
CMS = Centers for Medicare and Medicaid Services
EMR = electronic medical record
FP = False positive
OR = operating room
PHH = postoperative hemorrhage or hematoma
PPV = positive predictive value
PSI = Patient Safety Indicator
TP = true positive
VHA = Veterans Health Administration

itations, including whether they identify true events or preventable events. The current study focuses on PSI 9, ie, postoperative hemorrhage or hematoma (PHH), a component of the CMS-tracked composite measure. Although some blood loss is expected with most operations, this PSI is designed to capture bleeding after a surgical procedure that is presumably serious enough to require a subsequent reparative procedure. As such, the numerator requires both a diagnosis code of hemorrhage or hematoma and a procedure code for hemorrhage control or hematoma drainage.[5]

Like other PSIs, this indicator was developed using a consensus panel of clinical experts.[1] Additional work showed that hospital-level rates of this PSI were positively associated with rates of other PSIs representing postoperative complications,[6] and occurrence of this PSI was associated with excess hospitalization days, hospital costs, and in-hospital deaths.[7,8] However, relatively little is known about how well this indicator identifies true complications (ie, its criterion validity or agreement with medical record review). An earlier related indicator, "postprocedural hemorrhage or hematoma," from the Complications Screening Program had a moderate confirmation rate by chart review.[9] Complications Screening Program investigators also found frequently associated process of care problems, suggesting potential use of the current PSI as a quality of care measure.[10] Such use would be unwarranted if flagged cases do not identify true cases experiencing an event. Therefore, we examined the positive predictive value (PPV) of this indicator in the Veterans Health Administration (VHA). To better understand the type of events detected by this indicator and their potential preventability, we also examined circumstances surrounding this complication and associated risk factors.

METHODS

Study design

This was a retrospective cross-sectional study using VHA administrative and electronic medical record (EMR) data from fiscal year 2003 through 2007 (October 1, 2002 to September 30, 2007). We obtained Institutional Review Board approvals from the Bedford VA Medical Center and the VA Boston Healthcare System.

Data sources

We used hospital discharge information (ie, demographics, ICD-9-CM–coded diagnoses and procedures, and discharge status) from the VHA's National Patient Care Database Patient Treatment File.[11] Per earlier PSI work, we eliminated nonacute care (eg, long-term care).[7,12] We accessed VHA EMR data using VistaWeb, a program enabling centralized access to EMR data from all VHA facilities.[13]

PHH definition

The indicator is defined as "cases of hematoma or hemorrhage requiring a procedure per 1,000 surgical discharges with an operating room (OR) procedure."[6] The numerator requires both a secondary diagnosis code for hemorrhage or hematoma complicating a procedure and a procedure code for hemorrhage control or hematoma drainage. The denominator excludes discharges where the condition was present on admission, or the hemorrhage control or hematoma drainage procedure occurred before the first OR procedure, was the only OR procedure, or was part of the initial operative procedure (see Appendix 1, available online only, for the full PHH definition including ICD-9-CM codes).[5,6]

Study population

Hospital sampling

We applied the PSI software (v. 3.1a) to the inpatient database to obtain individual PSI counts and composite scores (ie, a combined measure that includes 11 PSIs).[14,15] From 158 acute care hospitals, we selected a representative sample of 28 hospitals based on individual PSI counts, composite rates, and geographic distribution (see Appendix 2, available online only, for sampling strategy and hospital characteristics). The observed PHH rate among sample hospitals was 3.8 per 1,000 (n = 614 cases), compared with a national VHA rate of 3.9 per 1,000 discharges at risk (n = 1,998 cases).

Case identification

We randomly selected 4 software-flagged cases of PHH per hospital. This total of 112 cases was based on power calculations using earlier reported PPVs and selected to ensure reasonably narrow confidence intervals (ie, 10% to 20%).[9]

Medical record abstraction

Two trained nurse-abstractors (KH, SM) conducted EMR reviews using a standardized data abstraction instrument

and guidelines adapted from AHRQ-developed preliminary tools.[16] The instrument included initial questions about demographics and ascertainment of the event; if a case was deemed a false positive (FP), abstraction ceased; for true positives (TPs), additional information was abstracted on risk factors, evaluation and management of the event, and patient outcomes.

To ensure consistency of abstracted information, we examined inter-rater reliability. The 2 nurses independently abstracted identical records in groups of 5 until they achieved ≥90% agreement across all questions, thereafter they abstracted different records. Study physicians (AB, HK) reviewed questions on which nurses disagreed, with resulting instrument revisions and/or guideline clarifications as appropriate. Study physicians also reviewed cases for clarification as required throughout the abstraction process. Additional inter-rater reliability assessment was performed on 5 charts toward the end of the abstraction process to check for abstractor reliability drift. Inter-rater reliability testing revealed 92% agreement (n = 58 questions) on the first 5 records and 93% on subsequent testing. Nurses agreed 100% on questions about event ascertainment (ie, identification of cases as TPs or FPs) on initial and subsequent inter-rater reliability testing ($\kappa = 1.0$).

Analyses

We categorized cases as TPs or FPs based on abstracted information and application of AHRQ's PHH definition. We calculated the PPV (ie, TPs/flagged cases) and associated 95% confidence intervals. Additionally, we examined FPs in detail to determine why they were flagged and gain insight into how the PSI might be improved. This included determining whether FPs resulted from inappropriate coding or limitations associated with coding (ie, the clinical intent of the indicator was not met despite correct coding) (see Appendix 1, available online only).

We compared selected demographics and assigned codes among TPs and FPs. For TPs, we examined clinical consequences of, and factors contributing to, the PHH occurrence, performing descriptive analyses of all variables.

To explore the relationship between processes of care and bleeding risk among TPs, we examined the association of bleeding risk with physician training status (resident versus attending) and procedure urgency (emergent versus nonemergent) using chi-square or *t*-tests as appropriate. Bleeding risk was defined by surgeon-estimated blood loss during the causative procedure, need for blood products intra- or postoperatively (yes/no), number of packed red blood cell units received during the hospitalization, or total number of units of all blood products received during the hospitalization. We also examined the association between bleeding risk and perioperative anticoagulant or antiplatelet use by defining mutually exclusive medication groups as follows: any receipt of IV heparin or warfarin, receipt of subcutaneous heparin (unfractionated or low molecular weight), receipt of antiplatelet agents, and no receipt of medication affecting hemostasis. We performed unadjusted logistic or linear regression as appropriate, with pairwise comparisons. We used SAS software, version 9.1 (SAS Institute Inc.) for analyses.

RESULTS

Of 112 flagged cases, 84 TPs met both coding and clinical criteria for a PPV of 75% (95% CI 66−83%). TP and FP patients were elderly (mean age 65 years or older), predominantly male (>96%) and white (>64%). Of the entire sample, 58% had a hematoma diagnosis code, 36% had a hemorrhage diagnosis code; and 6% had both. TPs were more likely to have a hematoma diagnosis or procedure code compared with FPs, although these differences were only significant for procedure codes (69% versus 50%; p = 0.18, and 80% versus 57%; p = 0.01, respectively) Thirty-six percent of FPs (n = 10) and 21% of TPs (n = 18) were discordant with respect to diagnosis and procedure codes, ie, they were assigned a hematoma diagnosis code and a hemorrhage procedure code or vice versa (not significantly different) (see Appendix 3, available online only, for TP and FP demographic and coding characteristics).

FP analysis

Four FPs (14%) failed to meet PHH coding criteria; in all 4 cases it was unclear why they received a 998.11 code (ie, Hemorrhage Complicating a Procedure code). Two of these had documented oozing during the index procedure without a discrete bleeding source. Although 3 of the 4 cases had appropriate procedure codes, these were for conditions unrelated to bleeding. Two cases had incision and drainage codes (86.04), one for a seroma, another for an abscess; the third case underwent exploratory laparotomy (code 54.12) for sepsis. These same codes are used for hematoma drainage.

An additional 24 cases did not fit the indicator's clinical intent. In 8 cases (29%), the postoperative complication was present on admission, resulting from a procedure performed before the index hospitalization; this included 5 hematomas and 3 hemorrhages. Six FPs (21%) had intraoperative bleeding controlled during the original procedure without subsequent bleeding or need for management. In 5 cases (18%), the postoperative hemorrhage or hematoma did not require a reparative procedure. In 3 other cases (11%), despite requiring management, the hematoma followed a noneligible procedure (ie, a diagnostic cardiac catheterization) (see Appendix 1, available online

Table 1. Hemorrhage-Hematoma Outcomes of True Positives (n = 84)

Outcome	
Return to OR for reparative/exploratory procedure, n (%)	69 (82.1)
Non-OR reparative procedure, n (%)*	16 (19.0)
Receipt of blood products during hospital stay, n (%)	54 (64.3)
RBCs, n (median, range)	51 (6U, 1–39U)
FFP, n (median, range)	30 (4U, 1–14U)
Platelets, n (median, range)	20 (2U, 1–7 U)
Cryoprecipitate, n (median, range)	5 (6U, 2–21U)
Lowest hematocrit, %, median (range)†	27.2 (12.2–42.2)
Moved to ICU, n (%)‡	11 (13.1)
Death, n (%)	6 (7.1)

*Acute airway compromise developed in 1 patient with a neck hematoma after carotid endarterectomy. The patient had a nonoperative hematoma evacuation at the bedside and then went to the OR for more definitive treatment.
†During the period after the index procedure until 7 days after discovery of the hemorrhage/hematoma or hospital discharge (whichever came first).
‡Does not include patients who were already admitted to the ICU for monitoring after the index procedure.
FFP, fresh frozen plasma; OR, operating room.

only, for explanation of eligible procedures). Two additional cases were present on admission and excluded for miscellaneous reasons: a patient with gastrointestinal bleeding post−bowel resection secondary to an arteriovenous malformation, and a patient admitted with an infected vascular graft that subsequently ruptured.

TP analysis

Outcomes

Sixty-nine TPs (82%) required at least 1 OR return to treat the hemorrhage or hematoma. Fifteen (18%) cases required a nonoperative procedure only, 11 of these were hematomas of the surgical wound treated with bedside incision and drainage. Nine TPs (11%) developed bleeding during the initial procedure and were treated intraoperatively, but then required an additional procedure for ongoing or recurrent hemorrhage or hematoma. Fifty-four TPs (64%) received blood products during the admission. Eleven TPs (12%) required ICU transfer because of bleeding consequences. Of 6 deaths (7%), 1 was directly attributed to bleeding (see Table 1).

Most hemorrhage/hematoma cases were discovered and addressed within 1 day of the index procedure (median 1 day; 25th, 75th percentile 0, 2 days). Up to 38% of cases were attributed to a possible technical problem occurring during the index procedure, such as suture unraveling, a vascular anastamosis defect, or localized vessel bleeding. The single most common physician-attributed bleeding cause was secondary to generalized oozing (n = 28; 33%); approximately one third (n = 10) of these also had an identified bleeding vessel. The cause was either unknown or not specifically documented in 42% of cases (see Fig. 1).

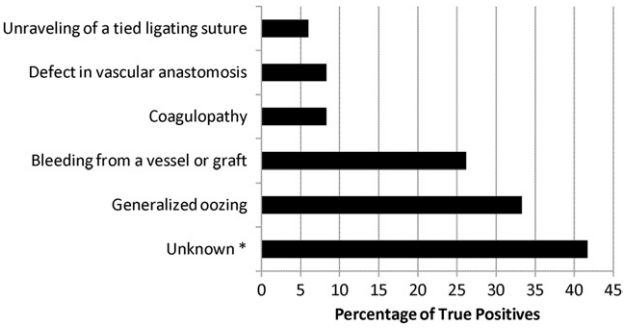


Figure 1. Physician-identified responsible factors for hemorrhage or hematoma. Multiple factors might be responsible for a given case. *Unknown, physician did not identify a cause for the bleeding.

Characteristics related to the index procedure

Of TP patients, 37% (n = 31) underwent an initial vascular procedure; 81% of these (n = 25) were associated with hematomas only. In contrast, only 42% of abdominal procedures (the next largest procedure category) were associated with hematomas only. Most index procedures (56%) were performed by a physician-in-training with documentation of attending supervision in >90%. Fourteen percent of procedures were emergent; 17% (n = 14) lost at least 1 L blood intraoperatively (see Table 2).

We found no substantial association between physician training status (resident versus attending) or procedure urgency and estimated blood loss during the index procedure, need for blood products during or after the procedure, or number of units received (data not shown, available from authors).

Coagulation-related factors

Only 6 TPs (7%) had medical conditions predisposing to bleeding present on admission (ie, primary hematologic disorder, advanced liver disease, or renal failure). In terms of potential medication-related coagulopathy, 35 TPs (43%) were on antiplatelets or anticoagulants before admission. Only 22% of patients on antiplatelets and none on anticoagulants had documentation that these were held before admission. Forty-three TPs (51%) received anticoagulants or antiplatelets in-hospital before the identified bleed (see Table 3).

Five patients on IV heparin had a partial thromboplastin time >150 seconds at some point perioperatively before the identified bleed. However, the physician attributed the bleeding specifically to coagulopathy in only 2 of these cases. Four of 9 patients on warfarin perioperatively had a therapeutic or supratherapeutic international normalized ratio (>2.0); however, only 1 of these patients who was also

Table 2. Characteristics of the Index Procedure

Variable	True positives (n = 84)
Anatomic region of index procedure, n (%)	
Vascular*	31 (36.9)
Abdominal/gastrointestinal†	19 (23.5)
Head and neck	8 (9.9)
Urologic	8 (9.9)
Cardiac‡	7 (8.3)
Skin and soft tissue§	5 (6.0)
Orthopaedic	4 (4.9)
Neurosurgical	1 (1.2)
Respiratory	1 (1.2)
Emergent or unscheduled procedure, n (%)	12 (14.3)
Rank of individual performing procedure, n (%)	
Resident or fellow	47 (55.6)
Attending	31 (36.9)
MD, unknown rank	5 (6.0)
Not documented	1 (1.2)
Estimated blood loss during initial OR procedure (per surgeon report), mL, median (range)	200 (0–4,000)
Receipt of RBCs during initial OR, n (%)	15 (17.9)
No. of RBC units, median (range)	4 (1–16)

*Fourteen lower-extremity procedures, 10 carotid procedures, 5 abdominal aortic aneurysm repairs, and 2 upper extremity procedures.
†Includes 1 splenectomy and 3 laparoscopic procedures.
‡Includes 3 minor cardiac procedures: 1 pacemaker, 1 implantable defibrillator placement, and 1 percutaneous coronary angioplasty.
§Two incision and drainages, and 1 each of breast reduction, facial rhytidectomy, and submental liposuction, and pedicle graft attachment.
OR, operating room.

on IV heparin with a partial thromboplastin time >200 seconds had coagulopathy identified as the bleeding cause. The perioperative platelet count was <100 in 18 TP patients and <50 in 3 patients. Of 30 patients with generalized oozing or coagulopathy identified as contributing to their bleeding, 9 were on an anticoagulant, 3 were on an antiplatelet (none were on clopidogrel), and 7 had a platelet count <100 (only 1 had a platelet count <50).

There was a trend toward higher intraoperative blood loss in patients on either an antiplatelet only or no medication affecting hemostasis versus those on anticoagulants, plus a trend toward greater need for blood products in patients receiving an antiplatelet versus the other medication groups. However, these differences were not significant in pairwise comparisons.

DISCUSSION

This is the first study to examine the criterion validity of the AHRQ PSI, PHH. We found that PHH has a moderate PPV (75%) when applied to VHA data. This estimate is consistent with related measures in the non-VHA setting and compares favorably with the upper range of reported PPVs of other recently validated PSIs representing postoperative complications (ie, 44% for postoperative venous thromboembolism to 83% for postoperative respiratory failure).[17,18] A precursor of the current PSI, which specified either a hemorrhage or hematoma secondary diagnosis code or a procedure code, was associated with a PPV of 78%.[9,19] Another study found a PPV of 75% comparing an ICD-9-CM diagnosis code of 998.1 (Hemorrhage or Hematoma Complicating a Procedure) to a clinical definition of postoperative hemorrhage after coronary artery bypass grafting (ie, return to surgery or blood product transfusion of ≥6 U).[20]

Table 3. Coagulation-Related Factors

Variable	True positives (n = 84)
Conditions present on admission, n (%)	
Coagulopathy (nonmedication-related)*	2 (2.4)
Advanced liver disease	2 (2.4)
End-stage renal disease	2 (2.4)
On anticoagulant or antiplatelet agent prior to admission, n (%)	37 (44.0)
Anticoagulant only (warfarin)	5 (6.0)
Antiplatelet only (aspirin or clopidogrel)	24 (28.6)
Anticoagulant and antiplatelet agent†	8 (9.5)
On anticoagulant or antiplatelet agent after admission and before bleed, n (%)‡	43 (51.2)
IV heparin or warfarin§	16 (19.0)
SC heparin (unfractionated or LMWH)	18 (22.2)
Antiplatelet agent only	9 (11.1)
Lowest perioperative temperature <96.8°F, n (%)‖	14 (16.7)
Lowest platelet count, $\times 10^3/mm^3$, median (range)¶	164 (4–531)
Highest PTT, seconds, median (range)¶	35.4 (21.8–240)
Highest INR, median (range)¶	1.28 (0.89–17.0)

*Two patients with possible idiopathic thrombocytopenic purpura versus myelodyplasia with preoperative platelet counts of 71 and 114.
†One of 8 patients on an anticoagulant and antiplatelet was on LMWH.
‡These were mutually exclusive groups (see text description). An additional 8 patients on anticoagulants were also on antiplatelets—3 in the IV heparin/warfarin group and 5 in the subcutaneous heparin group. Four patients in the antiplatelet only group were on both clopidrogel and aspirin; none were on only clopidogrel.
§Included 7 patients on IV heparin, 3 on warfarin, 6 on both. Patients who received only intraoperative heparin are not included.
‖Intraoperative temperatures were rarely available in the electronic medical record. We therefore used the time period starting at the index procedure until 24 hours postprocedure.
¶This was the most extreme value documented during the time period after the index procedure until discovery of the postoperative hemorrhage/hematoma.
INR, international normalized ratio; LMWH, low molecular weight heparin; PTT, partial thromboplastin; SC, subcutaneous.

Identification of true events—coding limitations and how to improve the indicator

Although 25% of cases were FPs, most were coded correctly but did not fulfill the indicator's clinical intent due to ICD-9-CM coding limitations. Our FP review suggests some basic modifications to the PSI algorithm and ICD-9-CM coding system could increase identification of flagged cases and the PPV.

With respect to cases resulting from a procedure performed before the current admission, although the indicator's PPV could be improved by adopting present on admission codes, which the VHA currently does not use, ignoring such cases could potentially result in loss of important data. The PSIs were intentionally designed to only use information from the index admission because of limitations of the dataset on which they were developed (ie, lack of linkage to information occurring outside the index admission).[1] Other researchers have advocated linking inpatient data to identify 30-day readmissions for PHH.[21] Because many procedures occur in the outpatient setting, outpatient linkage is also important. If we used a 30-day linkage to inpatient or outpatient data, 7 of the 8 present on admission cases would have been considered as TPs (the PPV would increase to 81%); we would have still identified 3 cases that were FPs (2 local surgical infections and the arteriovenous malformation case).

This indicator is intended to identify bleeding that was missed or not fully controlled during the initial procedure, as well as to identify postoperative bleeding serious enough to require a reparative procedure. Many of our FPs had intraoperative or postoperative bleeding but no subsequent reparative procedure, yet had procedure codes associated with the index procedure suggesting hemorrhage control (eg, 39.98, Control of Hemorrhage Not Otherwise Specified). Codes dealing with hemorrhage/hematoma diagnosis or control could be made more specific by associating a time factor with them (ie, intraoperative versus postoperative, or using a different code for bleeding control during the initial procedure versus a subsequent reparative procedure).

To decrease misidentification of cases that are not bleeds, the codes intended to represent hematoma drainage could be made more specific, for example, 86.04 (Other Incision And Drainage of Skin and Subcutaneous Tissue) or 54.12 (Reopening of Recent Laparotomy Site). The PPV impact of such enhancements, including data source linkage and coding changes, requires additional study because of potential trade-offs between sensitivity and specificity.

Additionally, we had 3 hematoma case requiring drainage associated with noneligible procedures (ie, diagnostic angiograms). However, angiography associated with an intervention such as angioplasty is eligible. Therefore, consideration should be given to including such diagnostic procedures.

Clinical implications/preventability

Although this indicator identifies clinically important events, because most TPs required at least one OR return and blood product transfusion, PHH and the other PSIs were also designed to identify complications resulting from quality of care problems. Theoretically, some PHH cases should be avoidable through appropriate perioperative care. Two studies of related indicators support this idea; the Complications Screening Program revealed potential quality problems in 37% of postprocedural hemorrhage or hematoma cases among Medicare patients[22]; a more recent pediatric study estimated that up to 57% of postoperative hemorrhage or hematoma cases were potentially preventable.[23] Because both studies used implicit review, the reproducibility of their findings is unclear.

Although not our main study aim, we abstracted data on risk factors, processes of care, and provider-identified causes to explore the preventability of events. Consistent with existing literature, we found that PHH was most commonly associated with vascular procedures.[21] This is not surprising considering the direct manipulation of large vessels and frequent concurrent use of systemic heparin. Preexisting nonmedication-acquired coagulopathies, which are known to be associated with increased bleeding risk, were rare among our sample.[24]

Up to 38% of procedures were associated with physician-documented potential technical problems. We investigated technical expertise and complication risk by examining trainee status and bleeding risk, but found no association; although we lacked adequate documentation on year of postgraduate training to be able to examine this further. The degree to which iatrogenic coagulopathy contributed to bleeding is unclear. Although medications affecting hemostasis must have played a role in some cases, fewer than half of TP patients received any anticoagulant or antiplatelet agent perioperatively, and few patients had abnormal clotting parameters or had coagulopathy explicitly cited as a bleeding cause. We also found no evidence of increased risk with respect to bleeding risk in patients receiving perioperative anticoagulants or antiplatelets.

Study limitations

Our study has a few limitations. This was a sample of male patients predominantly. However, patient sex should not affect physician documentation or complication coding. We cannot report on sensitivity, specificity, or negative predictive value; this was beyond the current study's scope and resources. Additionally, the ability to detect complications

depends not only on coding accuracy and completeness, but also on completeness of physician documentation. Tying sanctions to PSI rates, as CMS is doing, might impact both coding and medical record documentation. Although PHH is not among the individual CMS-tracked PSIs, it is part of the composite PSI. However, the VHA is less subject to financial incentives than other health care systems. Therefore, this should not have meaningfully affected findings.

Although several procedures were associated with potential technical problems and a few patients on anticoagulants had supratherapeutic clotting parameters, we cannot determine the actual number of cases associated with quality of care problems. This is due in part to lack of established evidence-based processes of care, possible physician documentation issues, the study's retrospective nature, the relatively small sample size, and lack of a control group. In addition, the nurse-abstractors were advised against drawing inferences about causation to maximize the reliability of findings. It was not our intent to make judgments about preventability.

Our findings are strengthened by the fact that cases were randomly drawn from a nationally representative sample of VHA hospitals. We assessed inter-rater reliability of abstraction at different time points and achieved a high level of abstractor agreement at each review. We thoroughly examined FPs and identified opportunities for improving this indicator's PPV.

CONCLUSIONS

In its current form, given its moderate PPV, it is reasonable to use PHH to screen for patient safety events. Additional indicator refinement is required using methods mentioned previously (eg, more specific diagnostic and procedure codes; data source linkage where feasible). Additional investigation about whether PHH identifies remediable quality of care problems is also necessary before it can be considered a definitive quality measure. This indicator is currently being validated in the non-VHA setting by AHRQ investigators. We await their results for comparison with our findings.

Author Contributions

Study conception and design: Borzecki, Kaafarani, Shin, Itani, Rosen

Acquisition of data: Borzecki, Kaafarani, Hickson, MacDonald

Analysis and interpretation of data: Borzecki, Kaafarani, Cevasco, Itani, Rosen

Drafting of manuscript: Borzecki, Rosen

Critical revision: Borzecki, Kaafarani, Cevasco, Hickson, MacDonald, Shin, Itani, Rosen

Final approval of the version to be submitted: Borzecki, Kaafarani, Cevasco, Hickson, MacDonald, Shin, Itani, Rosen

Acknowledgment: Special thanks to Patrick Romano, MD, MPH, principal investigator of the AHRQ QI Validation Project who served as a consultant on the study; Hillary Mull, MPP, for her technical assistance in developing the electronic abstraction form; Qi Chen, MPH and Shibei Zhao, MPH for programming assistance; Amresh Hanchate, PhD for methodologic advice; Dijon Fasoli, RN, PhD for conducting inter-rater reliability assessments; and Susan Loveland, MAT for programming expertise and assistance with identifying the hospital sample and flagged cases.

REFERENCES

1. McDonald KM, Romano PS, Geppert J, et al. Measures of patient safety based on hospital administrative data—the patient safety indicators. Technical review 5 (Prepared by the UCSF-Stanford Evidence-Based Practice Center under contract #290-97-0013). AHRQ Publication no. 02-0038. Rockville, MD: Agency for Health Care Research and Quality; August 2002.
2. National Quality Forum. NQF endorsed standards. Available at: http://www.qualityforum.org/Measures_List.aspx. Accessed June 1, 2009.
3. RHQDAPU Program overview: reporting hospital quality data for annual payment update. Available at: http://www.qualitynet.org/dcs/ContentServer?cid=1138115987129 &pagename=QnetPublic%2FPage%2FQnetTier2&c=Page. Accessed February 1, 2010.
4. Centers for Medicare and Medicaid Services. Fiscal year 2009 quality measure reporting for 2010 payment update. Available at: http://www.cms.hhs.gov/HospitalQualityInits/downloads/HospitalRHQDAPU200808.pdf. Accessed February 1, 2010.
5. Agency for Healthcare Research and Quality. Patient Safety Indicators: technical specifications. March 2003. Version 3.1. (March 12, 2007). Rockville, MD: Agency for Healthcare Research and Quality; 2007.
6. Agency for Healthcare Research and Quality. AHRQ quality indicators: guide to patient safety indicators. March 2003. Version 3.1 (March 12, 2007). Rockville, MD: Agency for Healthcare Research and Quality; 2007.
7. Rosen AK, Rivard P, Zhao S, et al. Evaluating the patient safety indicators: how well do they perform on Veterans Health Administration data? Med Care 2005;43:873–884.
8. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA 2003;290:1868–1874.
9. Lawthers AG, McCarthy EP, Davis RB, et al. Identification of in-hospital complications from claims data. Is it valid? Med Care 2000;38:785–795.
10. Iezzoni LI, Davis RB, Palmer RH, et al. Does the Complications Screening Program flag cases with process of care problems? Using explicit criteria to judge processes. Int J Qual Health Care 1999;11:107–118.
11. VA Information Resource Center. VIReC Research User Guide: FY2002 VHA Medical SAS Inpatient Datasets. Available at:

http://www.virec.research.med.va.gov/References/RUG/RUG-Inpatient02.pdf. Accessed December 1, 2009.

12. Rivard P, Elwy AR, Loveland S, et al. Applying Patient Safety Indicators (PSIs) across healthcare systems: achieving data comparability. Vol 2:7−25. Rockville, MD: Agency for Healthcare Research and Quality. AHRQ publication no. 005-0021-CD; 2005.
13. Veterans Health Administration. VistAWeb overview. Available at: http://vista.med.va.gov/vistaweb/. Accessed January 10, 2010.
14. Agency for Healthcare Research and Quality. Patient Safety Indicators Software SAS Documentation. March 2003. Version 3.2 (March 10, 2008). Rockville, MD: Agency for Healthcare Research and Quality; 2008.
15. Agency for Healthcare Research and Quality. AHRQ Quality Indicators: Patient Safety Indicator Composite SAS Documentation Version 3.1a (March, 2007). Rockville, MD: Agency for Healthcare Research and Quality; 2007.
16. Agency for Healthcare Research and Quality. AHRQ Quality Indicators Validation Pilot. Available at: http://qualityindicators.ahrq.gov/validationpilot.htm. Accessed March 1, 2010.
17. White RH, Sadeghi B, Tancredi DJ, et al. How valid is the ICD-9-CM based AHRQ patient safety indicator for postoperative venous thromboembolism? Med Care 2009;47:1237–1243.
18. Utter GH, Cuny J, Sama P, et al. Detection of postoperative respiratory failure: how predictive is the Agency for Healthcare Research and Quality's Patient Safety Indicator? J Am Coll Surg 2010;211:347−354e1−29.
19. Iezzoni LI, Foley SM, Heeren T, et al. A method for screening the quality of hospital care using administrative data: preliminary validation results. Qual Rev Bull 1992;18:361–371.
20. Hartz AJ, Kuhn EM. Comparing hospitals that perform coronary artery bypass surgery: the effect of outcome measures and data sources. Am J Public Health 1994;84:1609–1614.
21. Shufelt JL, Hannan EL, Gallagher BK. The postoperative hemorrhage and hematoma patient safety indicator and its risk factors. Am J Med Qual 2005;20:210–218.
22. Weingart SN, Iezzoni LI, Davis RB, et al. Use of administrative data to find substandard care: validation of the complications screening program. Med Care 2000;38:796–806.
23. Scanlon MC, Harris JM 2nd, Levy F, Sedman A. Evaluation of the agency for healthcare research and quality pediatric quality indicators. Pediatrics 2008;121:e1723–1731.
24. Christensen MC, Krapf S, Kempel A, von Heymann C. Costs of excessive postoperative hemorrhage in cardiac surgery. J Thorac Cardiovasc Surg 2009;138:687–693.
25. Rosen AK, Loveland SA, Romano PS, et al. Effects of resident duty hour reform on surgical and procedural patient safety indicators among hospitalized Veterans Health Administration and Medicare patients. Med Care 2009;47:723–731.

Appendix 1. Postoperative Hemorrhage or Hematoma: Detailed Definition and ICD-9-CM Codes

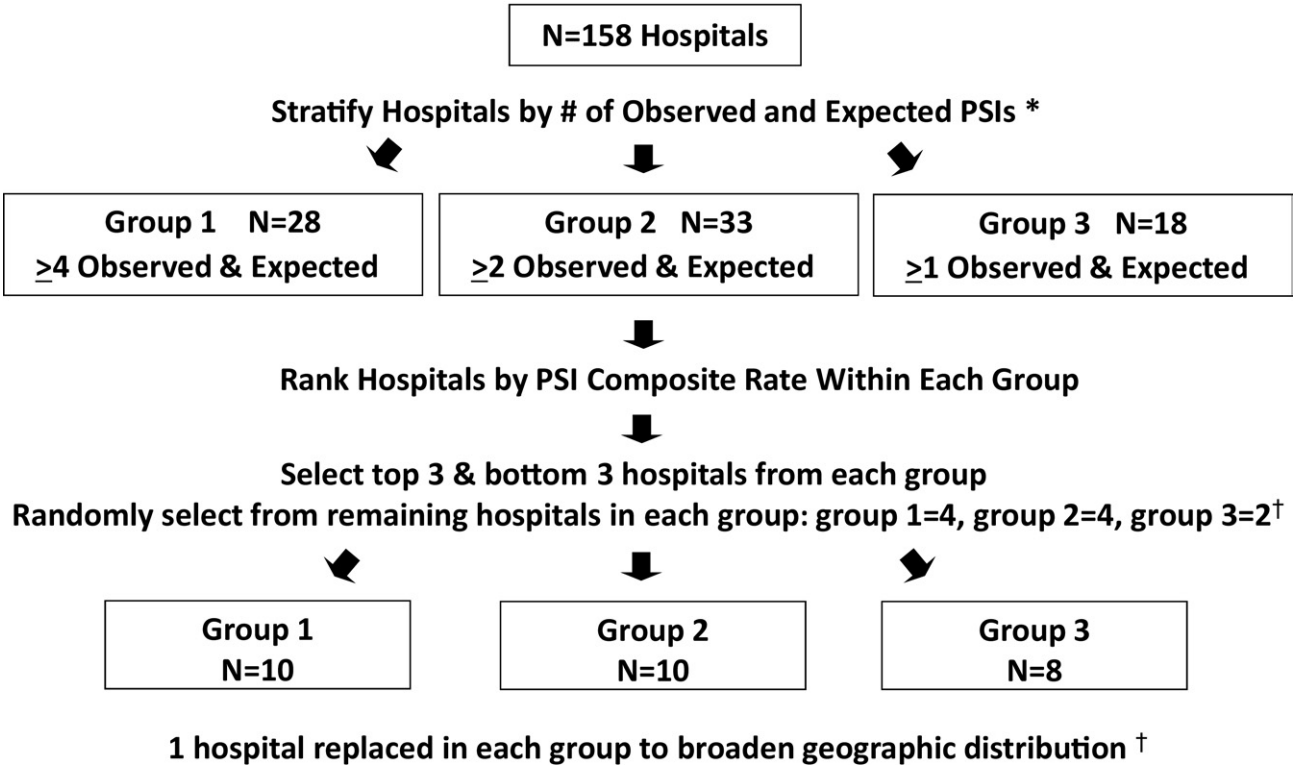
Definition	Cases of hematoma or hemorrhage requiring a procedure per 1,000 surgical discharges with an operating room procedure*	
Numerator	Discharges among cases meeting the inclusion and exclusion rules for the denominator with both of the following: ICD-9-CM code for postoperative hemorrhage (998.11) or postoperative hematoma (998.12) in any secondary diagnosis field† ICD-9-CM code for postoperative control of hemorrhage or for drainage of hematoma in any procedure code field	
Denominator	All surgical discharges 18 years and older defined by specific DRGs and an ICD9-CM code for an operating room procedure‡ Exclude cases: With preexisting condition (principal diagnosis or secondary diagnosis present on admission, if known) of postoperative hemorrhage or postoperative hematoma Where the only operating room procedure is postoperative control of hemorrhage or drainage of hematoma Where a procedure for postoperative control of hemorrhage or drainage of hematoma occurs before the first operating room procedure. MDC 14 (pregnancy, childbirth and the puerperium)	
Control of postoperative hemorrhage ICD-9-CM procedure codes	28.7	Control of Hemorrhage After Tonsillectomy and Adenoidectomy
	38.8x	Other Surgical Occlusion of Vessels (various sites)
	39.41	Control of Hemorrhage Following Vascular Surgery
	39.98	Control of Hemorrhage Not Otherwise Specified
	49.95	Control of (Postoperative) Hemorrhage of Anus
	57.93	Control of (Postoperative Hemorrhage of Bladder
	60.94	Control of (Postoperative) Hemorrhage of Prostate
Drainage of hematoma ICD-9-CM procedure codes	18.09	Other Incision of External Ear
	54.0	Incision of Abdominal Wall
	54.12	Reopening of Recent Laparotomy Site
	59.19	Other Incision of Perivesicle Tissue
	61.0	Incision and Drainage of Scrotum and Tunica and Vaginalis
	69.98	Other Operations on Supporting Structures of Uterus
	70.14	Other Vaginotomy
	71.09	Other Incision of Vulva and Perineum
	75.91	Evacuation of Obstetrical Incisional Hematoma of Perineum
	75.92	Evacuation of Other Hematoma of Vulva or Vagina
	86.04	Other Incision with Drainage of Skin and Subcutaneous Tissue

*Reported as rates, risk-adjusted for age, sex, and modified DRGs.

†Per the ICD-9-CM codebook, these codes are actually for hemorrhage or hematoma "complicating a procedure." Therefore, intraoperative bleeds would be coded identically.

‡See PSI technical specifications for the complete list of eligible surgical DRG and operating room (OR) procedure codes.[5] Eligible OR procedures include all procedures that are considered valid OR procedures by the DRG grouper; procedures such as a diagnostic cardiac catheterization are excluded, but percutaneous coronary angioplasty is included.

Appendix 2. Hospital Sampling Strategy



*The expected PSI count of a specific facility was calculated as the national VHA PSI rate multiplied by the facility's PSI denominator. Group 1 included facilities with at least four safety-related events in both the expected and observed numerators of each PSI; group 2 had at least two events; group 3 had at least 1 event.
†We randomly selected 4 remaining hospitals for groups 1 and 2, and 2 for group 3. We then replaced 1 of the chosen facilities in each group to improve geographic diversity. The final sample (n = 28) included hospitals representing 18 of the VHA's 21 regional health care networks; 32% were from the South, 29% from the West, 21% from the Northeast, and 18% were from the Midwest. Median number of hospital beds was 155 (range 62 to 360 beds). Eighty-nine percent were major or very major teaching hospitals.[25]

Appendix 3. Characteristics of Sample Patients

Variable	All flagged cases (n = 112)	True positives (n = 84)	False positives (n = 28)
Age, mean (SD)	66.8 (11.7)	67.5 (11.9)	64.5 (10.9)
Sex, male, n (%)	109 (97.3)	81 (96.4)	28 (100)
Race, n (%)			
White	77 (68.8)	54 (64.3)	22 (78.6)
Black	11 (9.8)	10 (11.9)	1 (3.6)
Hispanic	8 (7.1)	7 (8.6)	1 (3.2)
Other/missing	17 (15.2)	13 (15.5)	4 (14.2)
Codes, n (%)*			
Hemorrhage diagnosis	40 (35.7)	26 (30.9)	14 (51.6)
Hematoma diagnosis	65 (58.0)	52 (61.9)	13 (46.4)
Both diagnoses	7 (6.3)	6 (7.1)	1 (3.6)
Hemorrhage procedure	26 (23.2)	14 (19.0)	12 (42.9)†
Hematoma procedure	79 (70.5)	63 (73.4)	16 (57.1)
Both procedures	7 (5.6)	7 (7.1)	0

Percentages represent column percents.
*See Table 1 for a description of codes.
†Significant difference ($p < 0.05$) between true positives and false positives.

How Valid is the AHRQ Patient Safety Indicator "Postoperative Respiratory Failure"?

Ann M Borzecki, MD, MPH, Haytham MA Kaafarani, MD, MPH, Garth H Utter, MD, MSC, FACS, Patrick S Romano, MD, MPH, Marlena H Shin, JD, MPH, Qi Chen, MPH, Kamal MF Itani, MD, FACS, Amy K Rosen, PhD

BACKGROUND: The Agency for Healthcare Research and Quality Patient Safety Indicator postoperative respiratory failure (PRF) uses administrative data to screen for potentially preventable respiratory failure after elective surgery based on a respiratory failure diagnosis or an intubation or ventilation procedure code. Data on PRF accuracy in identifying true events is scant; a recent study using University HealthSystem Consortium data found a positive predictive value (PPV) of 83%. We examined the indicator's PPV in the Veterans Health Administration.

STUDY DESIGN: We applied the Patient Safety Indicator software (v.3.1a) to fiscal year 2003–2007 VA discharge data. Trained abstractors reviewed medical records of 112 software-flagged PRF cases. We calculated the PPV and examined false positives to determine reasons for incorrect identification and true positives to determine clinical consequences and potential risk factors of PRF.

RESULTS: Seventy-five cases were true positive (PPV 67%; 95% CI, 57–76%); 13% were identified by a diagnosis code, 53% by a procedure code, 33% by both. Of false positives, 19% represented coding errors, 76% represented nonelective admissions. Of true positives, 28% of patients died, 56% had an American Society of Anesthesiologists level higher than II. Of associated index procedures, 53% were abdominal/pelvic, and 56% lasted >3 hours.

CONCLUSIONS: Based on our and University HealthSystem Consortium's findings, PRF should continue to be used as a screen for potential patient-safety events. Its PPV could be substantially improved in the Veterans Health Administration through introduction of an admission status code. Many PRF-identified cases appeared to be at high risk, based on patient and procedure-related factors. The degree to which such cases are truly preventable events requires additional assessment. (J Am Coll Surg 2011;212:935–945. © 2011 by the American College of Surgeons)

Postoperative pulmonary complications are an important contributor to surgical risk, occurring in an estimated 6.8% of all surgical patients.[1] Guidelines exist for assessing perioperative pulmonary risk and recommend preventive strategies for high-risk patients.[2] Respiratory failure—usually defined as unplanned intubation or prolonged ventilation—is considered to be the most serious of the respiratory complications because of its high morbidity, mortality, and associated costs.[3]

Given the potential preventability, seriousness, and association with high resource use, respiratory failure is an attractive result with which to measure quality of surgical care. Postoperative respiratory failure (PRF), administrative data-based outcomes, has been developed by the Agency for Healthcare Research and Quality (AHRQ), as part of its Patient Safety Indicators (PSIs) module.[4] Because its reliance on administrative data, the AHRQ developers cautioned that PRF (PSI 11 and other PSIs) should be used as a screen for potential safety problems, highlighting areas in which quality of care should be investigated as opposed to being used as a definitive patient safety measure.[4] However, based on evidence of its validity to date, PRF was recently endorsed by the National Quality Forum as a hospital performance measure.[5-8] It also is a component of a com-

Disclosure Information: Nothing to disclose.

Supported by the Department of Veterans Affairs Health Services Research and Development service grant number SDR 07-002.

Abstract presented at AcademyHealth Annual Research Meeting, Boston, MA, June 2010.

Received July 31, 2010; Revised September 26, 2010; Accepted September 27, 2010.

From the Center for Health Quality, Outcomes and Economic Research, Bedford VAMC, Bedford (Borzecki), Department of Health Policy and Management, Boston University School of Public Health (Borzecki, Chen, Rosen), Boston University School of Medicine (Borzecki, Itani), Department of Surgery, Tufts Medical Center (Kaafarani), Department of Surgery, VA Boston Healthcare System (Itani), Harvard Medical School (Itani), Center for Organization, Leadership, and Management Research, VA Boston Healthcare System, Boston (Shin, Chen, Rosen), MA, Center for Healthcare Policy and Research (Utter, Romano), Department of Surgery (Utter), and Department of Internal Medicine (Romano), University of California, Davis, Sacramento, CA.

Correspondence address: Ann M Borzecki, MD, MPH, CHQOER, Bedford VA Hospital (152), 200 Springs Rd, Bedford, MA 01730. email: amb@bu.edu

ISSN 1072-7515/11/$36.00
doi:10.1016/j.jamcollsurg.2010.09.034

Abbreviations and Acronyms

AHRQ = Agency for Healthcare Research and Quality
EMR = electronic medical record
IRR = inter-rater reliability
OR = operating room
PPV = positive predictive value
PRF = postoperative respiratory failure
PSI = Patient Safety Indicator
UHC = University HealthSystem Consortium
VHA = Veterans Health Administration

posite PSI measure that the Centers for Medicare and Medicaid Services are tracking as part of their hospital reporting initiative.[9,10]

The PRF indicator is defined as cases of acute respiratory failure per 1,000 elective surgical discharges.[4] The numerator requires either an acute respiratory failure diagnosis code or an intubation or mechanical ventilation procedure code.[11] Recent risk-adjusted rates of this indicator in the nonfederal setting range from 5.3 to 15.8 events per 1,000 discharges, depending on the subgroup studied.[12] Previous work has demonstrated the face,[13] construct (through its association with related measures),[4] and predictive validity of this indicator (ie, its ability to predict an outcome such as death).[5,6] For example, hospital-level PRF rates were highly correlated with rates of other PSIs representing postoperative complications, and the occurrence of PRF was associated with >5-fold increased lengths of stay and costs, and >40-fold increased death rates in the Veterans Health Administration (VHA).[5,6] However, there are relatively few data on its criterion validity (ie, its agreement with an accepted standard, such as medical record information). To this end, we and other groups are conducting chart validation studies of PRF (as well as several of the other PSIs).

The purpose of this study is to build on a recent study performed in the nonfederal setting, using University HealthSystem Consortium (UHC) data. This study found a high positive predictive value (PPV) of 83% for the PRF indicator, signifying it has good accuracy for identifying true events.[7] However, this study also raised concerns that many PRF cases were likely not preventable.[7] Given potential differences with respect to patient case-mix and coding practices, it is important to examine if similar findings exist in other health care systems before conclusions can be made about the indicator's performance. Here we examined the validity of this indicator in the VHA, specifically its PPV. Additionally, we examined potential risk factors for and circumstances surrounding this complication to better understand the nature of events detected by this indicator and their potential preventability.

METHODS

This was a retrospective observational study using VHA administrative and electronic medical record (EMR) data from fiscal year 2003 through 2007 (October 1, 2002 to September 30, 2007). The study protocol was approved by the Bedford VA Medical Center and the VA Boston Healthcare System Institutional Review Boards.

Data sources

Data sources included the VHA's National Patient Care Database Patient Treatment File and the EMR. The Patient Treatment File contains administrative information on all VHA discharges, including demographics, diagnoses (principal and secondary ICD-9-CM codes), surgical and nonsurgical procedures (ICD-9-CM codes), and summary information (eg, admission/discharge dates, discharge status).[14] Because the PSIs were designed to screen for patient-safety events in nonfederal acute care hospitals, per previous work we eliminated the nonacute portion of care (eg, rehabilitation or long-term care), yielding a hospital discharge file containing only acute care.[6,15] We accessed the VHA's EMR data using VistAWeb, a program that allows centralized access to EMR data from all VHA facilities.[16]

PRF definition

Table 1 presents the full PRF definition and specific ICD-9-CM codes.[4,11] As shown, the numerator requires either a secondary diagnosis code for acute respiratory failure or an intubation or mechanical ventilation procedure code. The denominator excludes discharges when the condition was present on admission, or when a tracheostomy procedure was the only procedure or was performed before the first operation. It also excludes patients with major respiratory or circulatory disorders based on their principal diagnosis because of their higher risk for respiratory failure and lower likelihood of preventability.

The denominator also excludes nonelective hospitalizations, reflecting the logic that postoperative respiratory failure is less likely to be preventable "in patients admitted for non-elective surgeries, or urgent/emergent conditions."[15] In nonfederal administrative databases, admission type (eg, elective) serves as a proxy for surgery status because admission type is available but surgery status is not. However, the VHA Patient Treatment File lacks an admission type (or surgery status) field. Because of this, we previously developed an algorithm based on expert clinical opinion to distinguish between elective

Table 1. Postoperative Respiratory Failure Detailed Definition and ICD-9-CM Codes

Definition	Cases of acute respiratory failure per 1,000 elective surgical discharges with an OR procedure*
Numerator	Discharges among cases meeting the inclusion and exclusion rules for the denominator with ICD-9-CM codes for: Acute Respiratory Failure (518.81 or 518.84) in Any Secondary Diagnosis Field, or Reintubation or Mechanical Ventilation Procedure as Follows: 96.04 (Insertion of Endotracheal Tube) ≥1 Day after the Major OR Procedure Code 96.70 (Continuous Mechanical Ventilation of Unspecified Duration) or 97.71 (Continuous Mechanical Ventilation for <96 hours) ≥2 Days after the Major OR Procedure Code 96.72 (Continuous Mechanical Ventilation for ≥96 hours) ≥0 Days after the Major OR Procedure Code
Denominator	All elective surgical discharges aged 18 years and older defined by specific DRGs and an ICD-9-CM code for an OR procedure† Exclude cases: With pre-existing (principal diagnosis or secondary diagnosis present on admission, if known) acute respiratory failure With an ICD-9-CM diagnosis code of neuromuscular disorder Where a procedure for tracheostomy is the only OR procedure or tracheostomy occurs before the first OR procedure MDC 14 (pregnancy, childbirth, and puerperium) MDC 4 (diseases/disorders of respiratory system) MDC 5 (diseases/disorders of circulatory system)

*Reported as rates, risk-adjusted for age, sex, and modified diagnosis related group.
†Elective procedure defined by admit type. See PSI technical specifications for the complete list of eligible surgical DRG and OR procedure codes.[4,11] Eligible OR procedures include all procedures that are considered valid OR procedures by the DRG grouper; procedures such as a diagnostic cardiac catheterization are excluded, but a percutaneous coronary angioplasty is included.
OR, operating room; DRG, diagnosis related group; MDC, major diagnostic category; MDC is based on the principal diagnosis.

and nonelective hospitalizations using diagnosis related groups (DRGs) adapted from a nonelective DRG list originally used with California hospitalization data, admission date and time, and principal procedure date and time.[15] Cases with admissions or procedures occurring after hours or on weekends are classified as nonelective, as are DRGs, suggesting management of acute conditions (eg, trauma-related DRGs).

Study population

Hospital sampling

We applied the PSI software (v. 3.1a) to the VHA inpatient database to obtain individual PSI counts and composite scores (ie, a combined measure that includes 11 PSIs).[17,18] We selected hospitals to obtain a reasonably sized sample that would be representative of the diversity of VA hospitals. Starting with 158 acute care hospitals, we categorized facilities into 3 groups based on their observed and expected individual PSI counts, exclusive of PSI 5 (ie, Foreign Body Left During Procedure) and PSI 8 (ie, Postoperative Hip Fracture), which had low rates across most hospitals and obstetric PSIs. We ranked hospitals by PSI composite score and chose the top 3 and bottom 3 from each group. We then randomly selected from the remaining hospitals within each group to achieve a final sample of 28 hospitals (Fig. 1). The final sample included hospitals representing 18 of the VHA's 21 regional health care networks; 21% were from the Northeast, 32% the South, 18% the Midwest, and 29% from the West. Eighty-nine percent were major teaching hospitals (ie, resident-to-bed ratio >0.25).[19] Median number of hospital beds was 155 (range 62 to 360 beds).

Case identification

We used criteria developed by Rivard and colleagues to identify the eligible pool based on presumed elective admission status[15] and randomly selected 4 flagged cases of PRF from each of the 28 hospitals, yielding a total of 112 cases. This number was based on power calculations using earlier reported PPVs and selected to ensure reasonably narrow confidence intervals (ie, 10% to 20%).[20] Because 1 hospital had only 1 flagged case, we randomly selected 1 case each from 3 other hospitals in group 3 (Fig. 1).

Instrument development/chart abstraction

We modified a data abstraction instrument and guidelines from preliminary versions developed by AHRQ-funded investigators.[21] The instrument included questions about demographics, case ascertainment and exclusions, patient and procedure-related risk factors, evaluation, and management. (A revised version of a preliminary AHRQ tool with similar questions was used in the UHC study.) We added questions on perioperative fluid management and patient risk factors, with additional modifications occurring iteratively through clinician review, pre-pilot testing by 2 trained nurse-abstractors, and formal pilot testing with inter-rater reliability (IRR) assessment. For IRR measurement, the 2 nurse-abstractors independently reviewed identical records in groups of 5 until they achieved ≥90%

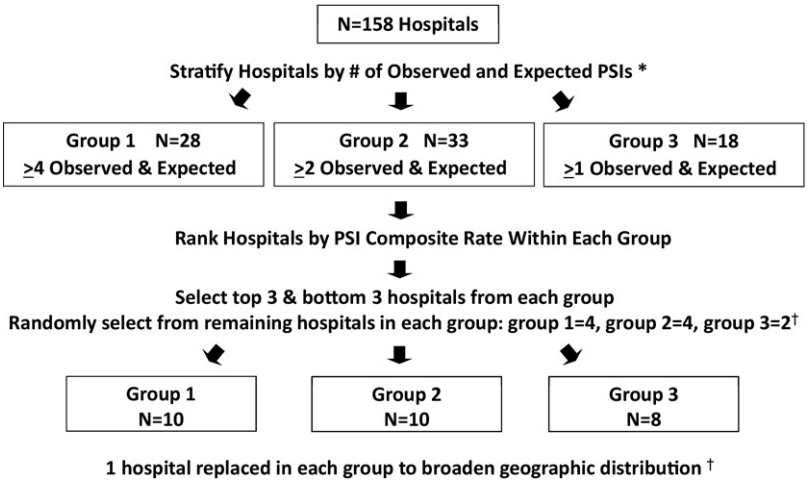


Figure 1. Hospital sampling strategy. *The expected Patient Safety Indicator (PSI) count of a specific facility was calculated as the national Veterans Health Administration PSI rate multiplied by the facility's PSI denominator. Group 1 included facilities with at least 4 safety-related events in both the expected and observed numerators of each PSI; group 2 had at least 2 events; group 3 had at least 1 event. †We randomly selected 4 remaining hospitals for groups 1 and 2, and 2 for group 3. We then randomly replaced 1 of the chosen facilities in each group to improve geographic diversity.

observed agreement across all questions, after which abstraction proceeded on different records. Abstraction disagreements were classified as being the result of an abstraction error (eg, the nurse abstracted an item incorrectly, such as noting the wrong value for weight or height) or an instrument problem (eg, a question was ambiguous, such as "what was the estimated blood loss during the procedure?" The response to which might vary based on whether the surgeon's, anesthetist's, or nurse's estimate was used). Nurse-abstractors corrected abstraction errors by reviewing the medical record together. They reviewed instrument-related issues with study clinicians (AB, HK), resulting in additional abstractor training, with instrument revisions and/or guideline clarifications as appropriate. Another IRR assessment was performed with 5 charts toward the end of the abstraction process to check for drift in abstractor reliability. Study physicians (AB, HK) also reviewed cases for clarification as necessary on an ongoing basis. This occurred with 6 cases; these tended to be cases in which abstractors could not determine why the patient remained ventilated postoperatively.

Once the initial criteria for inter-rater agreement were met, abstractors entered information directly into an electronic form hosted on a secure server on the VHA intranet, from which data were extracted into a relational database for subsequent analysis.

IRR testing revealed 90% agreement (n = 49 questions) on the initial 5 records and 91% on subsequent testing. Nurse-abstractors agreed 100% on questions pertaining to ascertainment of the event (ie, identification of cases as true positives [TPs] or false positives [FPs]) on initial and subsequent testing ($\kappa = 1.0$).

Analyses

Based on abstraction results and application of AHRQ's PRF definition, we categorized cases as TPs or FPs. Following UHC study methods, we calculated the PPV (ie, TPs/flagged cases) and associated 95% confidence intervals in 2 ways: based on whether the ICD-9-CM codes that resulted in the case being flagged were appropriate (coding criteria) and whether the case represented PRF from a clinical perspective (clinical criteria).[7] With respect to the second issue, we also classified FPs based on whether they failed to meet numerator or denominator criteria (Table 1). By classifying FPs as due to either coding inaccuracies or failure to meet the indicator's clinical intent, we hoped to gain insights into how to improve the PSI. Our characterization of cases differed slightly from the UHC study because we considered nonelective cases as FPs based on clinical criteria as opposed to coding criteria.

We compared TPs and FPs with respect to selected demographics and assigned codes using *t*-tests or chi-square as indicated. We also examined TPs to determine the clinical consequences of PRF and factors contributing to its occurrence, and performed descriptive analyses of relevant variables. Analyses were performed using SAS software, version 9.1 (SAS Institute Inc.).

Table 2. Characteristics of Sample Patients

Variable	All flagged cases (n = 112)	True positives (n = 75)	False positives (n = 37)
Age, y, mean (SD)	68.6 (10.2)	68.5 (9.9)	68.9 (10.9)
Sex, male, n (%)	111 (99.1)	74 (98.7)	22 (100)
Race, n (%)			
White	79 (70.5)	52 (69.3)	27 (73.0)
African American	15 (13.4)	11 (14.7)	4 (10.8)
Hispanic	5 (4.5)	4 (5.3)	1 (2.7)
Other/missing	13 (11.6)	8 (10.7)	5 (13.5)
ICD-9-CM codes, n (%)*			
Diagnosis code only	13 (11.6)	10 (13.3)	3 (8.1)
Procedure code only	65 (58.0)	40 (53.3)	25 (67.6)
Both diagnosis and procedure codes	34 (30.4)	25 (33.3)	9 (24.3)
Procedure codes, n (%)			
96.04	51 (45.5)	41 (54.7)	10 (27.0)†
96.70	1 (0.9)	0	1 (2.7)
96.71	10 (8.9)	6 (8.0)	4 (10.8)
96.72	37 (33.0)	18 (24.0)	19 (51.4)†

Percentages represent column percents. Cases were flagged per the Patient Safety Indicator algorithm.
*See Table 1 for a description of codes.
†Significant difference (ie, $p < 0.05$) between true positives and false positives.

RESULTS

Of 64,293 eligible hospitalizations among sample hospitals, 871 were flagged for PRF (ie, an observed rate of 13.5 per 1,000), compared with 2,850 flagged cases per 201,192 eligible hospitalizations nationally in the VHA (14.2 per 1,000). Of the 112 reviewed cases, 105 were TPs by coding criteria, for a PPV of 94% (95% CI 88–97%). When accounting for both coding and clinical criteria, the number of TP cases decreased to 75, for a PPV of 67% (95% CI 57–76%). Table 2 shows demographic and coding characteristics of all flagged cases, TPs, and FPs. The TP and FP groups were elderly (mean age 69 years) and predominantly male and white. Of the entire sample, 58% of cases were flagged based on a procedure code only, 12% were flagged based on a diagnosis code only, and 30% had both a diagnosis and procedure code. TP cases differed from FP cases with respect to assigned procedure codes; a 96.04 procedure code (Insertion of an Endotracheal Tube ≥1 Day after the Major Operating Room [OR] Procedure) was more common among TPs; a 96.72 procedure code (Continuous Mechanical Ventilation for ≥96 Hours 0 or More Days after the Major OR Procedure) was more common among FP cases (55% versus 27%, and 51% versus 24% respectively; $p < 0.05$).

FP analysis

Of the 7 patients (6%) that did not meet coding criteria, 2 remained on the ventilator immediately postoperatively for <96 hours; 4 had postoperative ventilation codes but no medical record documentation that they were on a ventilator; 1 patient was coded as if he had an OR procedure when he did not. An additional 30 patients did not fit clinical criteria; 28 of these did not satisfy denominator criteria because they were nonelective hospital admissions. Two additional patients did not fulfill the numerator criteria, as the patient was kept on the ventilator postoperatively for airway protection. Of the nonelective admissions, 5 would also have been excluded by other criteria, such as being intubated for respiratory failure before the first operation.

TP analysis

Outcomes

Median length of stay was 21 days (25^{th}, 75^{th} percentile 12, 36 days). Thirteen TP patients (17%) underwent tracheostomies; 2 TP patients were discharged with tracheostomies. No one was discharged on a ventilator. There were 21 deaths (28%).

Procedure-related risk factors

Of the TP cases, 53% (n = 40) involved abdominal/pelvic procedures; the next largest category was head and neck procedures at 16% (n = 12) (see Table 3 for additional details). Sixty-five TP patients (87%) had respiratory failure after the first OR procedure; in 6 (8%) this followed a second procedure and in 4 (5%) it followed a third procedure. Of the 10 TP patients (13%) who had respiratory failure after a nonelective procedure, in all but 1 this was a procedure resulting from a complication

Table 3. Procedure-Related Risk Factors

Variable	True positives (n = 75)
Anatomic region of preceding procedure, n (%)	
Head and neck*	12 (16.0)
Chest†	5 (6.7)
Abdomen/pelvis‡	40 (53.3)
Extremity§	9 (12.0)
Spine	
Cervical	3 (4.0)
Thoracolumbar	1 (1.3)
Percutaneous/endoscopic	3 (4.0)
Other‖	1 (1.3)
Nonelective procedure, n (%)¶	10 (13.3)
Type of anesthesia, n (%)	
General only	53 (70.7)
General + regional#	19 (25.3)
General + local (at incision)	2 (2.7)
Conscious sedation + local	0 (0)
Local only**	1 (1.3)
Epidural retained for postoperative pain control	15 (20.0)
OR blood loss, mL, median (range)	200 (0–5,100)
OR fluid input, mL, median (range)	2,950 (0–15,560)
Duration of operation, minutes, median, (range)	235 (35–795)
No. of OR procedures before respiratory failure (median, range)	1 (1–3)

Although the indicator is intended to exclude patients with principal diagnoses of circulatory or respiratory disorders, the 5 cases involving the chest and 6 vascular cases (5 carotid and 1 femoral anastomosis) all received an appropriate noncirculatory or respiratory-related principal diagnosis. (Patients with cerebrovascular disease get assigned to a major diagnostic category of "nervous system.")

*Includes 5 carotid vascular procedures.

†Includes 4 esophagectomies and 1 video-assisted thoracoscopic surgery.

‡Includes 1 splenectomy, 4 laparoscopic procedures, and 8 nephrectomies.

§Includes 1 femoral anastomosis revision; other procedures are orthopaedic.

‖Penile prosthetic implant.

¶Procedure status was obtained from the operative report.

#18 epidurals and 1 peripheral nerve block.

**One patient received only local anesthetic for a carotid angioplasty.

associated with the initial elective procedure. Fifty-three (71%) patients received general anesthesia only. Nineteen TP patients (25%) received regional anesthesia; 18 received an epidural, the majority of whom retained it for postoperative pain control, and 1 received a peripheral nerve block. Forty-two cases (56%) underwent an operation lasting >3 hours (see Table 3). Twenty percent lost at least 1 L blood intraoperatively. Thirty-one percent (n = 23) received at least 3 L IV fluid intraoperatively (crystalloid, colloid, and blood products combined; although this information was available for only 46 patients [61%]).

Table 4. Patient-Related Risk Factors

Variable	True positives (n = 75)
BMI, mean (SD)	27.7 (6.1)
Conditions present on admission	
Respiratory	
COPD	24 (32.0)
Current smoker	30 (40.0)
Obstructive sleep apnea	6 (8.0)
Cancer of lung or respiratory tract history	6 (8.0)
Asthma	2 (2.7)
Cardiac	
CHF history	8 (10.7)
Old MI (>30 days before admission)	8 (10.7)
SVT	9 (12.0)
Valvular heart disease	9 (12.0)
Neurologic	
CVA history	11 (14.7)
Alcohol abuse (within month before admission)	6 (8.0)
Alzheimer's disease	6 (8.0)
Quad-, para-, or hemiplegia	3 (4.0)
Other	
Diabetes	29 (38.7)
Chronic kidney disease	11 (14.7)
ASA class	
II	9 (12.0)
III	34 (45.3)
IV	8 (10.7)
Not documented	24 (32.0)
Preoperative albumin (g/dL), median (range)	3.7 (1.7–5.2)

ASA, American Society of Anesthesiologists; BMI, body mass index; CHF, congestive heart failure; SVT, supraventricular tachycardia.

Patient-related risk factors

Sixty-one percent (n = 46) had at least 1 pre-existing condition recognized as a risk factor for respiratory complications (eg, obstructive sleep apnea, COPD, smoking, alcohol use, or congestive heart failure), with COPD and smoking being most common at 32% and 40%, respectively.[2] Sixty percent had at least 1 risk factor for cardiac complications, with diabetes, and chronic kidney disease being the most common.[22] All but 8 TP patients (11%) had at least 1 substantial comorbidity (see Table 4 for additional details on individual comorbidities). The American Society of Anesthesiologists' class was higher than II in 82% of TP cases in which it was documented. Thirty-three percent were obese (ie, body mass index ≥30). Of 57 cases with a documented preoperative albumin level, it was <3.0 in 16% (n = 9).

Contributing factors

Seventy-nine percent of cases (n = 59) were initially extubated after the procedure but were reintubated later; 20%

(n = 15) remained intubated at the conclusion of the procedure (1 case of respiratory failure did not involve ventilation). Mean interval between the preceding operation and reintubation was 3.6 ± 5.2 days (median 2.0, 25th, 75th percentile 0, 4 days).

The most common provider-identified contributing factors included fluid overload (45%), pneumonia (43%), and atelectasis (31%), with these commonly co-occurring (see Fig. 2). Seventy-one percent had more than one contributing cause to respiratory failure. Eight cases represented primarily cardiac causes (ie, ischemia/infarction/atrial dysrhythmia with rapid ventricular response and congestive heart failure) as the cause of respiratory failure. Five cases represented cardiopulmonary arrests; of the 3 autopsies, 1 was a primary cardiac cause (ie, a massive myocardial infarction) and 2 were primarily respiratory (ie, a massive pulmonary embolism and aspiration). Upper airway causes of PRF were relatively rare (only 3 were due to upper airway edema); oversedation as the sole cause of PRF did not occur in any cases.

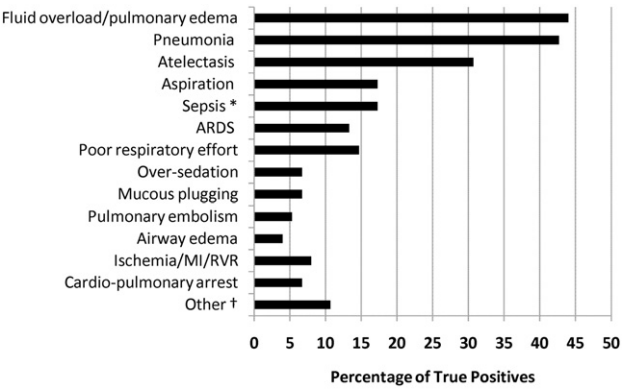


Figure 2. Physician-identified responsible factors for postoperative respiratory failure. Multiple factors can be responsible for a given case. *Sepsis unrelated to pulmonary process. †Includes COPD exacerbations or underlying COPD (n = 2), bronchospasm (n = 1), altered mental status (n = 1), thick secretions (n = 2). Ischemia/MI/RVR, myocardial ischemia/infarction/rapid ventricular response with congestive heart failure.

DISCUSSION

This is 1 of only 2 studies, to our knowledge, that examine the criterion validity of the AHRQ PSI, PRF in its current form, and the first to do so in the VHA.[7] We found that PRF has a modest PPV when applied to VHA data. Based on coding criteria, 94% of cases met its technical definition and 67% (95% CI 57–76%) met coding and clinical criteria, thereby representing true postoperative respiratory failure events. The majority of FPs represented patients with nonelective hospitalizations; incorrect coding was the next most common category. The in-hospital death rate for the TPs was markedly elevated (ie, 28%). PRF was most commonly associated with abdominal/pelvic procedures. A majority of the procedures were done under general anesthesia alone and were >3 hours long. Patient-related respiratory and cardiac comorbidities that can contribute to postoperative respiratory failure, such as COPD; current smoking; congestive heart failure history; and earlier MI, were common. Most cases represented reintubations after the initial operation, with volume overload, pneumonia, and atelectasis as the most common contributing conditions.

Identification of true cases

Comparison with other studies

The UHC study, involving 609 cases from 18 nonfederal academic medical centers, had a similar observed rate of PRF among sample hospitals at 14.7 per 1,000. The UHC investigators also found that 94% of cases met coding criteria, although a higher percentage, 83% (95% CI 77–89%), met both coding and clinical criteria.[7] As noted, our PPV results based on coding criteria are not directly comparable because the UHC data includes admission status codes (ie, FPs resulting from nonelective cases were considered as due to coding errors). The UHC study additionally had minor differences in classification of PRF cases based on investigators' interpretation of the indicator (ie, they considered patients with cardiac arrests as clinical FPs). Had we likewise categorized our patients, the PPV in this study would have dropped to 64%.

Comparison with older data on related measures reveals PPVs similar to our current findings both in the VHA and Medicare setting. We previously compared the original AHRQ PRF definition, defined solely by a diagnosis code for respiratory failure, with chart-based complication data from the National Surgical Quality Improvement Program.[23] The related National Surgical Quality Improvement Program clinical definition was "on the ventilator for more than 48 hours postoperatively." This work resulted in AHRQ-adopted revisions to the previous indicator. By capturing cases with either a diagnosis code for respiratory failure or intubation or ventilation procedure codes with associated timing specifications, the PSI's PPV decreased from 74% to 68%, and its sensitivity improved from 19% to 63%. The Complications Screening Program study, a sentinel patient-safety study using hospital discharge data from the early 1990s developed the indicator "Postoperative Pulmonary Compromise."[24] Although this indicator differed because it included no ventilation codes and only diagnostic codes for pulmonary congestion, pulmonary insufficiency, and acute pulmonary edema, its PPV was 72% (33 of 46) among Medicare beneficiaries.[4,20,24]

Coding-related strengths of indicator

Our current findings support the previous modifications to the AHRQ PRF indicator. Had we imposed the original PRF requirement of a diagnosis code of respiratory failure only, we would have detected just 46% of TPs. Consistent with UHC data, we found that it was rare for this indicator to identify cases of respiratory failure that were present on admission. None of our flagged cases had respiratory failure present on admission, compared with 0.8% of UHC cases.[7] Therefore, unlike some of the other PSIs, this indicator does a good job distinguishing complications or conditions that are present on admission from complications occurring during the hospital stay. Although the VHA does not code for present on admission status, in settings that have present on admission coding, reported rates have been similarly low, ranging from 1% using single institutional data[25] to just <7% using California and New York state data.[26]

Coding-related limitations of indicator

Because patients undergoing emergency procedures are at higher risk for nonpreventable events, they are intended to be excluded. Because there is currently no way to identify emergency procedures from administrative data, emergency admissions are considered a proxy for emergency procedures. The VHA has an additional limitation because admission status is not identified in the administrative database. Using a previously developed algorithm with face validity,[15] we incorrectly identified 28 flagged cases as elective admissions (25%). However, one might argue that PRF associated with nonelective admissions is still important to capture because some of these cases might be preventable. Had we retained cases classified as FPs based only on admission status, our PPV would have been 88% (95% CI 81–94%; ie, more comparable with the UHC findings).[7] Notably, even in UHC member hospitals that code data with respect to admission status, incorrect admission status coding was still the most frequent reason for FPs, occurring in 5% of cases.[7]

The indicator also excludes patients undergoing noninvasive ventilation (bilevel positive airway pressure/continuous positive airway pressure) unless they also have an acute respiratory failure diagnosis. In our sample, 1 TP patient received bilevel positive airway pressure postoperatively and remained on the ventilator prophylactically after a second surgery. This was flagged as PRF because of the ventilation associated with the second procedure, even though the clinical event was associated with the first. The effect on PPV of adding noninvasive ventilation (ie, procedure code 93.90) to the PRF definition requires additional study.

Additionally, the indicator makes no distinction between respiratory failure from a primarily pulmonary versus cardiac cause. One could argue that this distinction is of little clinical value because both cardiac and respiratory complications are common, potentially preventable, and serious, as they can result in respiratory failure, cardiopulmonary arrest, and death.

Apparent coding-related errors were seen in 6% of our flagged cases and 10% of UHC cases.[7] In the UHC study, after accounting for incorrect coding of hospitalization status, coding errors were most commonly due to a diagnosis code of acute respiratory failure not supported by medical record documentation; in our sample, coding errors were most commonly due to inappropriate coding for ventilation procedures. Although diagnostic terms can frequently be vague and subject to interpretation, procedure coding should be more straightforward. Although it might be difficult to determine the exact number of hours for which a patient was ventilated, it is unclear why FP patients who were not ventilated postoperatively received a mechanical ventilation code as if they were.

Are true cases preventable?

An important issue with respect to use of any potential quality measure is whether the identified adverse events are associated with process or systems failures and are therefore preventable. In theory, respiratory failure should be preventable through appropriate preoperative, intraoperative, and postoperative care. Additional support for this aspect of construct validity comes from the Complications Screening Program study; implicit physician review revealed potential quality problems in 27% of postoperative pulmonary compromise cases (n = 44) and only 2% of controls.[27]

As noted, guidelines exist for preoperative pulmonary (and cardiac) screening of patients undergoing elective procedures.[2,22] The goal is to determine their complication risk, manage modifiable risk factors preoperatively (eg, optimize lung/cardiac function through appropriate medication use in a patient with symptomatic COPD), and provide appropriate perioperative and postoperative management.

Although not our primary goal, we tried to get some idea of preventability by abstracting data on risk factors and processes of care. As in the UHC TP group, both patient- and procedural-related risk factors were very common. Compared with the UHC sample, VHA cases involved older patients with an especially high prevalence of COPD and smoking; despite this, essentially all TP patients in both samples had an American Society of Anesthesiologists class higher than I (among those for whom it was documented). Although the indicator excludes patients with a principal diagnosis of major respiratory or circulatory diseases, a large percentage of patients in both our and the UHC study had considerable respiratory or circulatory co-

morbidities impacting their risk. In both studies, TPs were commonly associated with procedures of at least moderate to high risk, of relatively long duration, and done under general anesthetic only.

The degree to which these risk factors could have been modified and the patient's risk decreased is uncertain. This is due in part to missing information in the EMR, as well as few existing strategies showing a benefit in either the preoperative, intraoperative, or postoperative setting.[28] With respect to preoperative management, although the majority of patients had at least documentation of a preoperative anesthesia assessment, because of resource limitations it was outside the scope of this study to determine whether patients had a change in preoperative comorbidity management or type of anesthesia used based on a preoperative assessment. Although volume overload was a common contributing cause, we found limited information with respect to intra- or postoperative fluid management (due to missing anesthesia reports or nursing flow sheets). In addition, information on use of postoperative preventive strategies (eg, lung expansion modalities) was frequently not documented in the EMR. Faced with similar findings because many of their cases appeared to be at high risk based on patient and procedural factors, the UHC investigators concluded that most of the identified cases were likely not preventable.[7]

Study strengths

Our cases were randomly drawn from a nationally representative sample of Veterans Affairs hospitals. Unlike the UHC study, we performed IRR testing of abstraction both early and late in the abstraction process and achieved high agreement between the registered nurses on abstracted data elements. In addition, our study clinicians were able to re-review cases when there were questions or concerns. In addition, unlike the UHC study, we also report on provider-identified circumstances surrounding the respiratory failure.

Study limitations

Similar to the UHC study, we are unable to report on sensitivity or specificity because we did not abstract cases that were not flagged by the PSI algorithm. This is because of difficulties in identifying an adequately representative sample that would not exceed time and staff resources available for chart abstraction. Our team does plan to examine PRF's negative predictive value among high-risk groups. Our study is also limited because our sample was relatively small, elderly, predominantly male, and had a high comorbidity burden. Despite this and the problems related to the VHA's lack of an admission status code, our findings with respect to risk factors were consistent with those of the UHC study, which had a larger population (507 TP patients) that was slightly younger and with more female patients (46% of TP patients). We cannot draw any conclusions as to the number of cases associated with potential quality of care problems or make recommendations to address these. This is partly a result of the retrospective nature of the study, relatively small sample size, lack of a control group, and missing data, such as anesthesia reports, which are often not available in VistAWeb.

How to improve the indicator

The PPV of the PRF indicator could be improved considerably in the VHA through introduction of an admission status code. This would also enable more accurate comparisons of VHA and non-VHA PRF rates. The UHC investigators have suggested that in non-VHA settings, misidentification of nonelective admissions could be improved by excluding cases that are suspicious for nonelective conditions based on principal diagnoses (eg, acute pancreatitis). Additional refinement would entail adding noninvasive ventilation codes. The UHC study additionally recommended excluding head and neck operations because they found several FP patients who were intubated for airway protection. Among our FP patients (including nonelective admissions), we had only 2 such patients. We also had 3 TP patients who were ventilated because of acute respiratory failure due to airway edema. Therefore, the trade-offs of any such modification will need to be carefully considered.

CONCLUSIONS

Based on both the UHC study and our findings, we think it is reasonable to use this indicator in its current form as a screen for postoperative respiratory failure events. Even with coding improvements that would enhance the identification of true complications, we have concerns about using this indicator for pay-for-performance. This is because of the lack of current evidence that such events are actually preventable through improvements of care. Additional study on whether identified events are avoidable is required, as is development of evidence-based methods to prevent such complications.

Author Contributions

Study conception and design: Borzecki, Kaafarani, Romano, Itani, Shin, Rosen

Acquisition of data: Borzecki, Kaafarani, Chen
Analysis and interpretation of data: Borzecki, Kaafarani, Utter, Romano, Chen, Itani, Rosen
Drafting of manuscript: Borzecki, Rosen
Critical revision: Borzecki, Kaafarani, Utter, Romano, Shin, Chen, Itani, Rosen

Acknowledgment: Special thanks to Kathleen Hickson, RN, MA and Sally MacDonald, RN for their assistance with abstraction instrument and guidelines modification, data abstraction, and data cleaning; Hillary Mull, MPP, for her technical assistance in developing the electronic abstraction form: Marisa Cevasco, MD, MPH for her clinical input; Amresh Hanchate, PhD for his methodologic advice; Dijon Fasoli, RN, PhD for conducting inter-rater reliability assessments; and Susan Loveland, MAT for programming expertise and assistance with identifying the hospital sample and flagged cases.

REFERENCES

1. Smetana GW, Lawrence VA, Cornell JE. Preoperative pulmonary risk stratification for noncardiothoracic surgery: systematic review for the American College of Physicians. Ann Intern Med 2006;144:581–595.
2. Qaseem A, Snow V, Fitterman N, et al. Risk assessment for and strategies to reduce perioperative pulmonary complications for patients undergoing noncardiothoracic surgery: a guideline from the American College of Physicians. Ann Intern Med 2006;144:575–580.
3. Johnson RG, Arozullah AM, Neumayer L, et al. Multivariable predictors of postoperative respiratory failure after general and vascular surgery: results from the patient safety in surgery study. J Am Coll Surg 2007;204:1188–1198.
4. Agency for Healthcare Research and Quality. AHRQ quality indicators: Guide to Patient Safety Indicators. Rockville, MD; March 2003; Version 3.1 (March 12, 2007). Rockville, MD: Agency for Healthcare Research and Quality; 2007.
5. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA 2003;290:1868–1874.
6. Rosen AK, Rivard P, Zhao S, et al. Evaluating the patient safety indicators: how well do they perform on Veterans Health Administration data? Med Care 2005;43:873–884.
7. Utter GH, Cuny J, Sama P, et al. Detection of postoperative respiratory failure: how predictive is the AHRQ Patient Safety Indicator? J Am Coll Surg 2010;211:347–354.
8. National Quality Forum. NQF endorsed standards. Available at: http://www.qualityforum.org/Measures_List.aspx. Accessed June 1, 2009.
9. RHQDAPU Program overview: reporting hospital quality data for annual payment update. Available at: http://www.qualitynet.org/dcs/ContentServer?cid=1138115987129&pagename=QnetPublic%2FPage%2FQnetTier2&c=Page. Accessed March 1, 2010.
10. Centers for Medicare and Medicaid Services. Fiscal year 2009 quality measure reporting for 2010 payment update. Available at: http://www.cms.hhs.gov/HospitalQualityInits/downloads/HospitalRHQDAPU200808.pdf. Accessed March 1, 2010.
11. Agency for Healthcare Research and Quality. Patient Safety Indicators: technical specifications. March 2003. Version 3.1 (March 12, 2007). Rockville, MD: Agency for Healthcare Research and Quality; 2007.
12. Agency for Healthcare Research and Quality. HCUPnet. National information on measures of health care quality based on the NIS, using the AHRQ Quality Indicators. Available at: http://hcupnet.ahrq.gov/HCUPnet.jsp?Id=6A4BE9A4C8EBA2BF&Form=SelPSIs1&JS=Y&Action=%3E%3ENext%3E%3E&_QITables=PSI11. Accessed July 21, 2010.
13. McDonald KM, Romano PS, Geppert J, et al. Measures of patient safety based on hospital administrative data—the patient safety indicators. Technical review 5 (Prepared by the UCSF-Stanford Evidence-Based Practice Center under contract no. 290-97-0013). AHRQ publication no. 02-0038. Rockville, MD: Agency for Health Care Research and Quality; August 2002.
14. VA Information Resource Center. VIReC Research User Guide: FY2002 VHA Medical SAS Inpatient Datasets. Available at: http://www.virec.research.va.gov/References/RUG/RUG-Inpatient02.pdf.
Accessed December 1, 2009.
15. Rivard P, Elwy AR, Loveland S, et al. Applying patient safety indicators (PSIs) across healthcare systems: achieving data comparability. Vol 2:7–25. Rockville, MD: Agency for Healthcare Research and Quality. AHRQ Publication No. 005-0021-CD; 2005.
16. Veterans Health Administration. VistAWeb Overview. Available at: http://vista.med.va.gov/vistaweb/. Accessed January 10, 2010.
17. Agency for Healthcare Research and Quality. Patient Safety Indicators software SAS documentation. March 2003, Version 3.2 (March 10, 2008). Rockville, MD: Agency for Healthcare Research and Quality; 2008.
18. Agency for Healthcare Research and Quality. AHRQ Quality Indicators: Patient Safety Indicator Composite SAS documentation. Version 3.1a (March, 2007). Rockville, MD: Agency for Healthcare Research and Quality; 2007.
19. Rosen AK, Loveland SA, Romano PS, et al. Effects of resident duty hour reform on surgical and procedural patient safety indicators among hospitalized Veterans Health Administration and Medicare patients. Med Care 2009;47:723–731.
20. Lawthers AG, McCarthy EP, Davis RB, et al. Identification of in-hospital complications from claims data. Is it valid? Med Care 2000;38:785–795.
21. AHRQ Quality Indicators Validation Pilot. Available at: http://qualityindicators.ahrq.gov/validationpilot.htm. Accessed March 1, 2010.
22. Fleisher LA, Beckman JA, Brown KA, et al. ACC/AHA 2007 guidelines on perioperative cardiovascular evaluation and care for noncardiac surgery: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2007;116:e418–e499.
23. Romano PS, Mull HJ, Rivard PE, et al. Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. Health Serv Res 2009;44:182–204.
24. Iezzoni LI, Foley SM, Heeren T, et al. A method for screening the quality of hospital care using administrative data:

preliminary validation results. Qual Rev Bull 1992;18:361–371.

25. Bahl V, Thompson MA, Kau TY, et al. Do the AHRQ patient safety indicators flag conditions that are present at the time of hospital admission? Med Care 2008;46:516–522.

26. Houchens RL, Elixhauser A, Romano PS. How often are potential patient safety events present on admission? Jt Comm J Qual Patient Saf 2008;34:154–163.

27. Weingart SN, Iezzoni LI, Davis RB, et al. Use of administrative data to find substandard care: validation of the complications screening program. Med Care 2000;38:796–806.

28. Lawrence VA, Cornell JE, Smetana GW. Strategies to reduce postoperative pulmonary complications after noncardiothoracic surgery: systematic review for the American College of Physicians. Ann Intern Med 2006;144:596–608.

Journal of Thrombosis and Haemostasis, **9**: 441–449 DOI: 10.1111/j.1538-7836.2010.04179.x

REVIEW ARTICLE

Future directions of stroke prevention in atrial fibrillation: the potential impact of novel anticoagulants and stroke risk stratification

K. P. CABRAL,* J. ANSELL† and E. M. HYLEK‡

**Department of Pharmacy, Maine Medical Center, Portland, ME; †Department of Medicine, Lenox Hill Hospital, New York, NY; and ‡Department of Medicine, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA*

To cite this article: Cabral KP, Ansell J, Hylek EM. Future directions of stroke prevention in atrial fibrillation: the potential impact of novel anticoagulants and stroke risk stratification. *J Thromb Haemost* 2011; **9**: 441–9.

Summary. Stroke prevention in atrial fibrillation is of paramount importance given its associated morbidity and mortality. The many challenges of warfarin limit its effective use in real-world clinical practice. We are entering an exciting therapeutic era as new classes of anticoagulants, including direct thrombin inhibitors, factor Xa inhibitors and novel vitamin K antagonists, are being evaluated for possible use in this patient population. If proven to be as efficacious as warfarin and safer, expanded use of these novel agents to lower risk subgroups may be justified. It is imperative that providers be aware of the many advantages and potential challenges posed by use of these novel agents in routine clinical care. An understanding of individual pharmacokinetic profiles and potential drug-drug and drug-disease interactions will translate into improved effectiveness in real-world practice.

Keywords: anticoagulation, atrial fibrillation, direct thrombin inhibitors, factor Xa inhibitors, novel anticoagulants, stroke prevention.

Introduction

Atrial fibrillation (AF) is the most common significant cardiac rhythm disorder in the world. In the United States alone, it currently affects an estimated 2.5 million people a year and is estimated to increase to approximately 8 million individuals by the year 2020 [1,2]. The increasing incidence is due to multifactorial causes and is attributed largely to the aging of the population, the rates of hypertension, obesity, diabetes mellitus and heart failure and increased survival with chronic cardiovascular disease. The morbidity and mortality associated with AF are substantial. Atrial fibrillation-related strokes are associated with a 30-day mortality of 24% [3]. The Framingham Heart Study has shown that the attributable risk of stroke for patients with AF increases from 1.5% at age 50–59 years to 23.5% at age 80–89 years [4]. It is projected that by the year 2050, more than half of the patient population with atrial fibrillation will be older than 80 years of age [5]. Older individuals are also at the highest risk of serious bleeding from anticoagulant therapy, highlighting the critical balance between risk and benefit.

Currently, the vitamin K antagonist, warfarin, is the mainstay of anticoagulant therapy for patients with atrial fibrillation [6]. Developed over 60 years ago, warfarin has been the most effective treatment to prevent ischemic stroke in patients with atrial fibrillation, decreasing the stroke risk by two-thirds when compared with placebo [7]. Despite its efficacy, warfarin remains under-used in clinical practice [8,9]. In fact, only 50–60% of those eligible AF patients are anticoagulated with warfarin, due to its high risk profile and complex pharmacokinetics and pharmacodynamics [10]. Maintaining its narrow therapeutic range in a real-world patient population is challenging due to warfarin's variable dose response, diet and medication interactions, and frequent monitoring. Analyses from clinical trial data have identified a minimum threshold of time in the therapeutic range (TTR) of at least 60% to maximize the benefit of warfarin [11,12]. In addition, warfarin has a slow onset of action, a long duration of action and a long elimination half-life, which makes it difficult to manage peri-procedural interventions. Due to these dosing complexities, warfarin is associated with a high rate of adverse events, which creates barriers to more widespread use [13,14].

Despite 60 years of experience with warfarin for stroke prevention in atrial fibrillation, many questions remain to be answered. Current investigation of novel anticoagulants, the mechanisms underlying thrombosis and propagation of atrial fibrillation, the role of genetics, stroke risk stratification, and the comparative effectiveness of different treatment strategies

Correspondence: Katherine P. Cabral, Department of Pharmacy, Maine Medical Center, 22 Bramhall Street, Portland, ME, USA.
Tel.: +1 207 662 2219; Fax: + 1 207 662 6619.
E-mail: cabrak@mmc.org

Received 27 October 2010, accepted 10 December 2010

across diverse subgroups, will further our understanding and inform optimal management of this enlarging patient population. In this review, we will provide an update and overview of promising new anticoagulant therapies and highlight areas of current uncertainty in the prevention of stroke in atrial fibrillation.

Novel oral anticoagulants

There are currently over 20 novel anticoagulants in development and more than half of them are oral agents. The new oral anticoagulants primarily focus on inhibiting a specific factor in the coagulation pathway by directly binding to the factor. Various binding mechanisms are employed depending on which factor is the target, including factors (Fs) II, V, VII, IX, X and XI. The agents that are most advanced in clinical research and closest to approval for various indications, are the direct thrombin (FIIa) inhibitors and the direct FXa inhibitors [15–18].

Direct thrombin (FIIa) inhibitors

Oral, direct thrombin inhibitors (DTIs) are synthetic, small molecules that interact directly with thrombin, blocking both free and clot bound thrombin [16]. Blocking directly at FIIa is a desirable strategy because it is the final step in blood coagulation. Thrombin plays a central role as a procoagulant by converting fibrinogen to fibrin, activating other substrates (FV, FVIII, FXI and FXIII) and activating platelet protease activated receptors [19]. It is also the most potent agonist of platelet aggregation. By binding directly to the active site of thrombin and minimally to plasma proteins, DTIs produce a predictable anticoagulant response [16,19].

The first oral DTI, ximelagatran, developed and studied over 5 years ago, was a prodrug of melagatran. It had predictable pharmacokinetics and anticoagulant response, required no monitoring and had minimal drug or food interactions. Ximelagatran was the first oral DTI to undergo clinical testing, showing that it was at least as effective as warfarin for stroke prevention in atrial fibrillation [20,21]. It was ultimately removed from the market due to its association with hepatic toxicity. Nevertheless, ximelagatran established a cautious precedent for inhibition of thrombin as a viable single target for anticoagulants.

Dabigatran etexilate

Dabigatran etexilate is the oral DTI that has been most extensively studied to date and is currently approved for prevention of venous thromboembolism (VTE) after total hip or knee replacement surgery in Europe and Canada. The US Food and Drug Administration recently approved the use of dabigatran for stroke prevention in AF [22,23]. Dabigatran etexilate is a prodrug that is rapidly converted to its active form, dabigatran. Dabigatran etexilate is a twice daily oral medication, with an onset of action of about 2 h. It has an elimination half-life of 12–17 h, which may be prolonged in renal insufficiency, as it is primarily eliminated through renal routes [24,25]. Dabigatran does not require routine monitoring and may avoid major cytochrome (CYP) P450 drug interactions

Table 1 Pharmacokinetic and pharmacodynamic characteristics of new agents and stage of development [16,24–32,65]

	Dabigatran	AZD0837	Rivaroxaban	Apixaban	Edoxaban	Betrixaban	YM-150	Tecarfarin
Target	Factor IIa	Factor IIa	Factor Xa	Factor Xa	Factor Xa	Factor Xa	Factor Xa	VKOR
Prodrug	Yes	Yes	No	No	No	No	No	No
Tmax (h)	1.5–3	0.7–1.5	2–4	1–3	1–2	NR	NR	3–5
VD (L)	60–70	NR	50	Reported as low	NR	NR	NR	NR
Half-life (h)*	12–17	9	9–13	9–14	9–11	19	NR	119
Metabolism	Conjugation	CYP3A4, 2C9, 2C19	CYP3A4, 2J2	CYP3A4	CYP3A4	NR	NR	Esterase pathways
Elimination	80% renal	NR	66% renal	25% renal	35% renal	Unchanged in bile	NR	NR
Administration/ dosing in AF	Fixed twice daily	Fixed once (ER) or twice daily (IR)	Fixed once daily	Fixed twice daily	Fixed once daily	Fixed once daily	Fixed once or twice daily	Monitored, once daily
Drug interactions	Potent P-gp inducers/ inhibitors	Potent CYP3A4 inhibitors	Potent CYP3A4 inhibitors & P-gp inducers/ inhibitors	Potent CYP3A4 inhibitors & P-gp inducers/ inhibitors	CYP3A4 inhibitors & P-gp inducers/ inhibitors	Low potential reported	NR	Expected to be minimal
Afib Trial stage of development	Completed phase III/ FDA approved	Completed phase II	Completed, phase III	Active, phase III	Active, phase III	Completed phase II	Completed phase II	Completed phase II

*In normal renal function. NR, not reported to date in literature; CYP, cytochrome P450.

because it is metabolized through conjugation [23–32]. (Table 1) It is, however, a P-glycoprotein (P-gp) substrate and implications of potential drug interactions will be discussed later in the paper. Dabigatran has recently been evaluated in a number of phase III clinical trials, including VTE prevention after orthopedic surgery, VTE treatment, and stroke prevention in AF [33–38]. Subsequent to a phase II evaluation in 502 patients with atrial fibrillation, two doses of dabigatran were chosen to be compared with warfarin in the Randomized Evaluation of Long-term anticoagulation therapY (RE-LY) trial [37]. This multicenter, prospective, open-label, randomized trial included patients with non-valvular atrial fibrillation at moderate to high risk of stroke or systemic embolism [38,39]. Patients were excluded for a number of reasons, but most notably, a history of heart valve disorders, stroke within 14 days or severe stroke within 6 months, conditions associated with an increased risk of bleeding, severe renal impairment (Cockcroft Gault creatinine clearance (Clcr) ≤30 mL min^{-1}) or active liver disease. The primary objective was to demonstrate dabigatran as non-inferior to warfarin in preventing stroke and systemic embolism. Over 18 000 patients worldwide were randomized to dabigatran etexilate 110 mg twice daily, dabigatran etexilate 150 mg twice daily or warfarin (target international normalized ratio (INR) goal 2.0–3.0).

The mean age of patients enrolled in RE-LY was 71 years and approximately two-thirds were men [38]. Stroke risk was assessed using the $CHADS_2$ risk scoring scheme, assigning 1 point each for congestive heart failure, hypertension, age ≥75 years and diabetes mellitus, and two points for history of prior stroke or transient ischemic attack (TIA) [40]. The mean $CHADS_2$ score was 2.1, with roughly equal proportions distributed among the respective $CHADS_2$ categories: 0–1, 2 and 3–6 [38]. The median duration of follow-up was 2 years, and 99.9% of patients achieved complete follow-up of at least 1 year. The mean TTR for the warfarin group was 64%. Dabigatran 110 mg twice daily was found to be non-inferior to warfarin in preventing stroke or systemic embolism (1.54% vs. 1.71% per year, $P = 0.30$) and superior in terms of major bleeding (2.87% vs. 3.57% per year, $P = 0.003$). Dabigatran 150 mg twice daily was found to be superior to warfarin in preventing stroke or systemic embolism (1.11% vs. 1.71% per year, $P < 0.001$) and non-inferior to warfarin in terms of major bleeding ($P = 0.32$). Both dabigatran doses resulted in fewer intracranial hemorrhages compared with warfarin ($P < 0.001$), and there were no signals of hepatic toxicity. Gastrointestinal bleeding was significantly increased with the dabigatran 150 mg dose and discontinuation rates were higher in both dabigatran groups compared with warfarin at 1 and 2 years. Consistent with other dabigatran phase III clinical trials, the most common reason for discontinuation was gastrointestinal symptoms, particularly dyspepsia (11.8%, 110 mg dose; 11.3%, 150 mg dose; 5.8%- warfarin; $P < 0.001$ for both). A higher rate of clinical myocardial infarction was observed with both doses, but neither was significant compared with warfarin (0.82 for the 110 mg dose, 0.81 for the 150 mg dose, vs. 0.64 for warfarin; $P = 0.09$ and 0.12, respectively), and overall vascular mortality was reduced.

Although results of this trial were impressive, several unanswered questions remain, warranting further investigation in this patient population, particularly regarding optimal dose selection. The US FDA recently approved the 150 mg twice daily dose and did not approve the 110 mg twice daily dose. A 75 mg twice daily dose was also approved based on pharmacokinetic extrapolations, to be used in patients with renal insufficiency. A recent analysis of RE-LY reassuringly showed similar results in patients with previous TIA or stroke for secondary prevention [41]. RELY-ABLE, a long-term follow-up study of RE-LY trial participants, will look at outcomes up to 28 months. It is estimated to be completed in late 2011 [42].

AZD0837

In addition to dabigatran etexilate, AZD0837 is another oral DTI in development that has shown promise in phase II testing in the AF population. AZD0837, a prodrug that is rapidly converted to its active form, has an average elimination half-life of 9 h and is eliminated through both the renal and hepatic routes [16]. It is currently being developed and studied as both a twice daily immediate-release regimen and a once daily extended-release formulation [43]. AZD0837 has undergone two phase II trials, one with each of the formulations in development. The first phase II trial was a randomized, parallel, dose-guiding study that compared AZD0837 150 mg twice daily or 350 mg twice daily with dose-adjusted warfarin (INR 2.0–3.0) [43]. Patients with atrial fibrillation at moderate risk, with one additional risk factor for stroke, were studied over 3 months at 20 sites within Europe. Results showed that the 150 mg dose had a similar safety profile as warfarin and was more tolerable than the 350 mg dose. Few major bleeding events occurred, with no difference in overall frequency between groups, but a higher rate of minimal bleeding events occurred with the 350 mg dose. More patients discontinued treatment in the 350 mg dose group, primarily due to gastrointestinal adverse effects. There was a low incidence of elevated liver enzymes in this population; however, a 10% increase in serum creatinine with both AZD0837 doses was noted. The elevations were transient, reversed after discontinuation and were not related to a decrease in glomerular filtration rate (GFR) [43]. The other completed phase II, dose-guiding study, investigated AZD0837 150, 300 and 450 mg once daily extended-release formula, and a 200 mg twice daily regimen, compared with dose-adjusted warfarin (INR 2.0–3.0) [44]. Results indicated that overall bleeding events were less common in the AZD 150, 300 and 200 mg groups compared with warfarin and similar to warfarin in the 450 mg group. The proportion of patients with adverse effects was similar between the AZD0837 and warfarin groups; however, the AZD0837 groups had higher discontinuation rates compared with warfarin. Similar to the earlier phase II trial, the most

common adverse effects associated with AZD0837 were gastrointestinal disorders. These adverse effects were less frequent in the lower AZD0837 doses. There was a slight increase in liver enzymes in the AZD0837 groups, and an increase of approximately 10% in serum creatinine, as noted in the earlier phase II trial.

Factor Xa inhibitors

Oral, FXa inhibitors bind directly to the active site of FXa and block the interaction with its substrate [16]. Factor Xa is strategically located at the convergence of the intrinsic and extrinsic pathways, effectively inhibiting thrombin generation from both sources. In addition, by not blocking thrombin directly, traces of thrombin may be able to escape neutralization, facilitating hemostasis and potentially providing an increased safety profile [16,19]. Furthermore, it may be advantageous to block more proximally in the coagulation cascade, thereby minimizing the amplification of thrombin generation that occurs downstream [16,19].

Rivaroxaban

Rivaroxaban is an oral FXa inhibitor currently in use in Europe and Canada for VTE prevention in the orthopedic population [45–49]. It is a potent, selective and reversible FXa inhibitor that inhibits free FXa both in solution and within a clot [16]. Rivaroxaban has a time to peak concentration of about 3 h and is partially metabolized through the CYP450 system. Two-thirds of the drug is renally eliminated, with a half-life of 9–13 h, and it does not require routine monitoring. It has been evaluated in phase III clinical trials for VTE prophylaxis in the orthopedic population, for VTE treatment, and for stroke prevention in AF [42,50].

Results of ROCKET-AF (Rivaroxaban once daily Oral direct factor Xa inhibition Compared with vitamin K antagonism for prevention of stroke and Embolism Trial in Atrial Fibrillation) were recently presented at the annual scientific sessions of the American Heart Association [51]. ROCKET-AF is a randomized, double-blind, double-dummy study with over 14 000 participants worldwide. The study compared rivaroxaban 20 mg once daily, or 15 mg once daily for Clcr 30–49 mL min^{-1}, with warfarin (target INR 2.0–3.0). A distinct difference between ROCKET-AF and previous trials is the medical complexity of the study population. Fifty-five percent of participants had a history of stroke, 62% had heart failure, and 87% had a $CHADS_2$ score of 3 or greater. Rivaroxaban was found to be non-inferior to warfarin in this high-risk population for prevention of stroke and systemic embolism. It did not achieve superiority in the intention-to-treat analysis, 2.12% for rivaroxaban vs. 2.42% for warfarin (hazard ratio (HR), 0.88; 95% confidence interval (CI), 0.74 to 1.03). Importantly, rates of hemorrhagic stroke were decreased with rivaroxaban, 0.26% vs. 0.44% (HR, 0.58; 95% CI, 0.38–0.89). Overall rates of major bleeding were similar, 3.60% vs. 3.45% (HR, 1.04; 95% CI, 0.90–1.20). Of note, the TTR achieved in this trial was 57.8% [50,51].

Apixaban

Apixaban, another oral FXa inhibitor in development, is currently emerging from phase III testing in orthopedic surgery, with ongoing trials for several indications including VTE treatment and atrial fibrillation [52,53]. Apixaban is administered twice daily and has an elimination half-life of 8–15 h [16,19,24]. It is eliminated through various pathways, with only 25% through the kidneys, suggesting an advantage for those with renal impairment. Although apixaban is metabolized through CYP3A4 to several metabolites, apixaban is not thought to directly inhibit or induce CYP450 and therefore has a low potential for major drug interactions [16]. Apixaban is currently being studied in two trials for stroke prevention in atrial fibrillation. ARISTOTLE (Apixaban for the Prevention of Stroke in Subjects With Atrial Fibrillation) will randomize over 18 000 patients with non-valvular AF to apixaban 5 mg twice daily or dose-adjusted warfarin (INR 2.0–3.0) [42,54]. This double-blind, parallel arm, event-driven, non-inferiority trial is expected to be completed in 2011. The Apixaban Versus Acetylsalicylic Acid to Prevent Stroke in Atrial Fibrillation Patients Who Have Failed or Are Unsuitable for Vitamin K Antagonist Treatment (AVERROES) trial was recently stopped early due to evidence of clear benefits of apixaban compared with aspirin among individuals with atrial fibrillation who either failed or were unsuitable for warfarin [52]. AVERROES included 5600 patients randomized to receive apixaban 5 mg twice daily or aspirin 81–324 mg once daily. Preliminary data were presented at the European Society of Cardiology Congress 2010 [55,56]. The study showed that apixaban led to a statistically significant reduction in the risk of stroke or systemic embolic events (relative risk (RR), 0.46; 95% CI, 0.33–0.64) and a reduction in cardiovascular hospitalizations, with no statistically significant increase in bleeding rates (RR, 1.14; 95% CI, 0.74–1.75). Although these results, compared with aspirin, are not unexpected, the AVERROES trial provides evidence that patients deemed to be at too great a risk for warfarin can be safely treated with apixaban.

Edoxaban

Edoxaban (DU-176b) is an orally active, small-molecule, reversible FXa inhibitor. It is rapidly absorbed, reaching a maximum concentration about 1–2 h after oral administration. Edoxaban has an elimination half-life of 9–11 h and 35% is renally excreted [16]. Favorable results have been reported from phase II studies of edoxaban for VTE prophylaxis in patients after total knee arthroplasty and total hip replacement [57,58]. In addition, edoxaban has recently completed phase II testing in atrial fibrillation. This study of over 1100 patients demonstrated a safety and tolerability profile similar to that of warfarin [59]. The doses from this trial, 30 and 60 mg once daily, are now being studied in a phase III, randomized,

double-blind, double-dummy trial, ENGAGE AF-TIMI 48 (Effective aNticoaGulation with factor xA next Generation in Atrial Fibrillation). An estimated 20 500 patients are to be enrolled from 1400 clinical sites worldwide. The planned treatment duration is 24 months and the study expects to conclude in early 2012 [42].

Betrixaban and YM150

Betrixaban and YM150 are two oral FXa inhibitors that have recently completed phase II testing in the AF population; however, neither agent has entered phase III trials yet. Betrixaban has a rapid onset of action and an elimination half-life of 19 h, allowing once daily dosing. It offers several potential advantages over other anticoagulants, including biliary elimination, therefore completely avoiding renal clearance. It has a low potential for major drug interactions because it is not a substrate for CYP450 [16,60–62]. In addition, the agent is being developed in conjunction with an intravenous antidote to neutralize the anticoagulant effects of betrixaban, as well as those of other FXa inhibitors [63].

Betrixaban recently completed EXPLORE-Xa, a dose-finding phase II trial in AF patients with one additional risk factor [64]. Three doses, 40, 60 and 80 mg, were compared with warfarin. Patients were an average of 74 years of age and had a mean $CHADS_2$ score of 2.2. There were fewer incidences of major and clinically relevant non-major bleeding with the 40 mg dose compared with warfarin (one vs. four patients). The 60 and 80 mg doses had bleeding rates similar to warfarin. No differences in liver function enzymes were noted. The only adverse effect that was more common with betrixaban compared with warfarin was gastrointestinal complaints, particularly with the two higher doses.

YM150 has minimal pharmacokinetic or pharmacodynamic data available to date. However, it is reported to be well tolerated in phase I studies with predictable pharmacokinetics in healthy young and elderly volunteers [16]. YM150 has undergone phase II, dose-finding studies in VTE prevention, as well as in atrial fibrillation. A phase II study in patients with non-valvular AF has been completed, in which YM150 was compared with warfarin. Patients are being recruited for a second phase II study, in which YM150 will be compared with warfarin in non-valvular AF patients, with a primary outcome of major bleeding [42]. This study, OPAL-2, is expected to be completed in late 2010.

Novel vitamin K antagonist: tecarfarin

The novel oral anticoagulants currently being studied in atrial fibrillation are not exclusively direct thrombin or FXa inhibitors. A new vitamin K antagonist in development, tecarfarin, is a selective, non-competitive inhibitor of vitamin K epoxide reductase (VKOR). It is developed to be similar to warfarin but is metabolized by esterases instead of the CYP450 system [65]. With a more predictable pharmacological profile and its potential lack of drug interactions, it is assumed to have clinical, real-world advantages over warfarin. Tecarfarin was evaluated in an open-label, phase II study of 66 AF patients to determine the safety, tolerability and effect of tecarfarin on TTR [66]. The TTR with tecarfarin was high; only 10% of patients had a TTR of $<45\%$. The mean interpolated TTR was 71.4% over the initial 3 weeks of tecarfarin therapy, resulting in a median daily dose of 15.6 mg per patient to maintain an INR of 2–3. This TTR is higher than typically observed with warfarin in clinical trials; however, patients were monitored intensively, with INRs drawn at least two to three times per week for the first 3 weeks. Although circumventing CYP450 metabolism is a true scientific advance, tecarfarin will still be subject to dietary vitamin K fluctuation and variability secondary to the variants in the VKOR complex subunit 1 (*VKORC1*) gene. In the absence of a warfarin control, more clinical data are needed on this agent.

Anticipated issues outside of the clinical trial setting

Risk stratification in diverse patient populations

Clinical trial participants are often younger with overall fewer chronic medical conditions compared with patients seen in routine practice. Low- to high-risk $CHADS_2$ scores range from 0 to 6 [40]. Trial participants in recent trials have had a mean $CHADS_2$ score of approximately 2.1 for RE-LY and 3.48 for ROCKET-AF. This is important not only from a stroke risk perspective, but also from a bleeding perspective, as bleeding risk has been shown to increase with $CHADS_2$ score [67–69]. The efficacy and safety of these novel agents will need to be closely monitored in the older and more medically complex patient populations.

Conversely, as newer agents with potentially improved safety profiles emerge and those barriers we currently face to the effective use of warfarin dissipate, the expansion of anticoagulant drugs to the lowest risk patients may be justified. The estimated yearly stroke risk among individuals with a $CHADS_2$ score of 0 is approximately 2% [40]. The recently published European Society of Cardiology guidelines for the management of atrial fibrillation incorporate risk modifiers (female gender, age 65–74 years, vascular disease) to consider among patients with a CHADS2 score of 0 or 1 to comprise a $CHADS_2VASc$ score. Anticoagulant therapy is recommended if two or more of these modifiers are present and anticoagulant therapy or aspirin if only one is present. If the risk of major hemorrhage is reduced with novel agents, then the threshold for their use will probably also be reduced, providing they are cost effective.

Reversibility and anticoagulant half-life

The oral FIIa and FXa inhibitors offer many advantages over warfarin, but also have potential limitations, one of which is the absence of a reversal agent. Their short half-lives, compared with warfarin, provide assurance that drug concentrations will decline relatively rapidly when therapy is discontinued in

patients with normal renal function. Nevertheless, reversibility in emergent situations, such as trauma, life-threatening bleeding, emergent surgery, or in patients with severe renal insufficiency, is unclear. Limited animal studies have shown that the anticoagulant effects of dabigatran and rivaroxaban may be reversed by infusions of recombinant FVIIa or prothrombin complex concentrates [71,72]. Of particular interest is the development of a specific modified FXa protein, lacking catalytic and membrane binding activities, that can competitively neutralize a FXa inhibitor [63,73]. The clinical value of all of these approaches remains to be determined in future clinical trials. In addition, the development of a reliable and valid laboratory assay to measure residual anticoagulant effects would be desirable.

With regard to the shorter half-lives of these drugs, adherence will be a key issue for effective use in clinical practice [74]. Respective of the faster onset of action, thrombotic risk associated with interruption is likely to be less of an issue, but post-procedural bleeding may increase if hemostasis is not fully achieved prior to resumption of these drugs. For dabigatran, it is recommended to discontinue the drug 'one to 2 days prior to invasive or surgical procedures for patients with a Clcr $\geq$50 mL min^{-1} or 3–5 days if Clcr $<$ 50 mL min^{-1}' [25]. It is also recommended to perhaps discontinue it 'earlier than 5 days in those patients undergoing major surgery or requiring complete hemostasis' [25]. Such recommendations must be guided by further clinical experience.

Table 2 Common P-gp transporters (list is not all inclusive) [79–81]

P-gp substrates	Atorvastatin
	Cyclosporine
	Digoxin
	Loperamide
	Quinidine
	Indinavir, nelfinavir, ritonavir
	Dexamethasone, hydrocortisone
	Vinblastine, vincristine
	Daunorubicin, doxorubicin, etoposide
P-gp inhibitors	Amiodarone
	Ceftriaxone
	Clarithromycin, erythromycin
	Cyclosporine
	Diltiazem
	Dipyridamole
	Hydrocortisone
	Ketoconazole, itraconazole
	Nicardipine, nifedipine
	Propranolol
	Quinine
	Quinidine
	Ritonair, saquinavir, nelfinavir
	Tamoxifen
	Tacrolimus
	Verapamil
P-gp inducers	Rifampin
	Clotrimazole
	Phenytoin
	Phenobarbital
	St. John's Wort

Drug interactions

An additional potential advantage that the emerging FIIa and FXa inhibitors may possess over warfarin is a decreased number of drug interactions. Warfarin's numerous food and drug interactions often lead to non-therapeutic INRs and decreased TTR. Newer agents have few, if any, interactions with food. However, they are not completely free of drug interactions. From the available data, almost all of the agents, excluding dabigatran, tecarfarin and potentially betrixaban, are metabolized to some degree by the CYP450 enzyme system. As numerous medications are metabolized via this pathway, it could be the source of potential medication interactions. Furthermore, the structure, location and activity of CYP450, particularly 3A4, overlap with the P-glycoprotein (P-gp) transport system, which may lead to additional drug interactions [75]. P-gp is an intracellular drug transport system that plays an important role, particularly in drug absorption and distribution [76]. The P-gp system is widespread, expressed on the surfaces of the gastrointestinal tract, brain, liver, kidney and capillaries, and acts as an efflux pump, preventing the uptake of foreign substances [76]. The activity of this extensive transport system is controlled in part through genetic factors; however, foods and medications can also influence its activity. Medications that inhibit P-gp will increase absorption of a substrate, thus increase its serum concentrations (area under the curve [AUC]), and medications that induce P-gp will have the opposite effects, decreasing serum concentrations [75] (Table 2). Because of its pharmacokinetic importance and potential implications in drug interactions, P-gp transport screening is often incorporated into the drug discovery process [76].

Dabigatran etexilate, rivaroxaban, apixaban and edoxaban are P-gp substrates [15,77]. Current data show that dabigatran yielded no clinically relevant interaction when given concurrently with the P-gp substrates, atorvastatin and digoxin. However, it resulted in significant decreases in AUC when given with the P-gp inducer, rifampin, and therefore the combination should be avoided in patients [22,77]. When given with the P-gp inhibitors, such as ketoconazole, verapamil, clarithromycin, quinidine and amiodarone, the AUC of dabigatran increased; however, it did not result in changes in the extent of absorption, maximum concentration, time to peak concentration or half-life. Therefore, if these particular agents are given with dabigatran, no dose adjustments are recommended at this time [22,77]. (Table 1) However, increased monitoring of renal function should be employed if amiodarone is given with dabigatran. The use of azole antifungals or HIV-protease inhibitors is a contraindication for rivaroxaban, and most likely also for apixaban. Most of these drugs are strong inhibitors of both CYP3A4 and P-glycoprotein, resulting in significant increases in AUC and changes in pharmacodynamic effects of the anticoagulant [77]. Additionally, the

concomitant use of a potent P-gp inhibitor is a protocol-specified exclusion criterion in the ENGAGE AF-TIMI 48 trial and the need for verapamil or quinidine mandates a protocol-directed 50% reduction of the edoxaban dose [78]. Increases in AUC concentrations of the novel anticoagulants from potential P-gp drug interactions may increase the risk of hemorrhage and therefore warrant caution. Phase IV registries will help to elucidate the potential hazards of these drug-drug and drug-disease interactions in real-world practice.

Conclusion

Stroke prevention in atrial fibrillation is entering an exciting therapeutic era, with new classes of targeted anticoagulants that avoid the many pitfalls of the current vitamin K antagonists. It is hoped that these promising new drugs will be at least as efficacious as warfarin and safer. Their ease of management will improve patients' quality of life and alleviate the workload of healthcare providers. All of these attributes may encourage providers to treat more patients with atrial fibrillation, many of whom are currently considered unsuitable for warfarin.

Disclosure of conflict of interests

K.P. Cabral states that she has no conflict of interest. J. Ansell is a consultant for Boehringer Ingelheim, Ortho McNeil, and Bristol Myers Squibb. E.M. Hylek is on the Advisory Board of Boehringer-Ingelheim, Bristol-Myers Squibb, Daiichi Sankyo, Merck, Pfizer and Sanofi-Aventis; and research for Bristol-Myers Squibb and Ortho McNeil. None of these achieve the level of 'significant'.

References

1 Miyasaka Y, Barnes ME, Gersh BJ, Cha SS, Bailey KR, Abrayaratna WP, Seward JB, Tsang TS. Secular Trends in Incidence of Atrial Fibrillation in Olmstead County, Minnesota, 1980-2000, and Implications on the Projections for Future Prevalence. *Circulation* 2006; **114**: 119–25.

2 Singer DE, Albers GW, Dalen JE, Fang MC, Go AS, Halperin JL, Lip GYH, Manning WJ. Antithrombotic therapy in atrial fibrillation: ACCP evidence-based clinical practice guidelines. *Chest* 2008; **133**: 546S–92S.

3 Hylek EM, Go AS, Chang Y, Jensvold NG, Henault LE, Selby JV, Singer DE. Effect of Intensity of Oral Anticoagulation on Stroke Severity and Mortality in Atrial Fibrillation. *N Engl J Med* 2003; **349**: 1019–26.

4 Lin H, Wolf PA, Kelly-Hayes M, Beiser AS, Kase CS, Benjamin EJ, D'Agostino RB. Stroke severity in atrial fibrillation: The Framingham Study. *Stroke* 1996; **27**: 1760–4.

5 Go AS, Hylek EM, Phillips KA, Henault LE, Selby JV, Singer DE. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention; the AnTicoagulantion and Risk Factors in Atrial Fibrillation (ATRIA) Study. *JAMA* 2001; **285**: 2370–5.

6 Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G. The pharmacology and management of the vitamin K antagonists: ACCP evidence-based clinical practice guidelines. *Chest* 2008; **133**: 160S–98S.

7 Hart RG, Benavente O, McBride R, Pearce LA. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. *Ann Intern Med* 1999; **131**: 492–501.

8 Waldo AL, Becker RC, Tapson VF, Colgan KJ. Hospitalized Patients With Atrial Fibrillation and a High Risk of Stroke Are Not Being Provided With Adequate Anticoagulation. *J Am Coll Cardiol* 2005; **46**: 1729–36.

9 Ogilvie IM, Newton N, Welner SA, Cowell W, Lip GYH. Under use of oral anticoagulants in atrial fibrillation; A systematic review. *Amer J Med* 2010; **123**: 638–45.

10 Connolly SJ, Eikelboom J, O'Donnell M, Pogue J, Yusuf S. Challenges of Establishing New Antithrombotic Therapies in Atrial Fibrillation. *Circulation* 2007; **116**: 449–55.

11 White HD, Gruber M, Feyzi J, Kaatz S, Tse HF, Husted S, Albers GW. Comparison of outcomes among patients randomized to warfarin therapy according to anticoagulant control: results from SPORTIF III and V. *Arch Intern Med* 2007; **167**: 239–45.

12 Connolly SJ, Pogue J, Eikelboom J, Flaker G, Commerford P, Franzosi MG, Healey JS, Yusuf S. Benefit of oral anticoagulant over antiplatelet therapy in atrial fibrillation depends on the quality of international normalized ratio control achieved by centers and countries as measured by time in therapeutic range. *Circulation* 2008; **118**: 2029–37.

13 Budnitz DS, Pollock DA, Weidenbach KN, Mendelsohn AB, Schroeder TJ, Annest JL. National surveillance of emergency department visits for outpatient adverse drug events. *JAMA* 2006; **296**: 1858–66.

14 http://www.usp.org/pdf/EN/patientSafety/drugSafetyReview2003-12-08.pdf; Accessed 23 February 2011.

15 Weitz JI, Hirsch J, Samama MM. New Antithrombotic Drugs. *Chest* 2008; **133**: 234S–56S.

16 Eriksson BI, Quinlan DJ, Weitz JI. Comparative Pharmacodynamics and Pharmacokinetics of Oral Direct Thrombin and Factor Xa Inhibitors in Development. *Clin Pharmacokinet* 2009; **48**: 1–22.

17 Agnelli G, Gallus A, Goldhaber SZ, Haas S, Huisman MV, Hull RD, Kakkar AK, Misselwitz F, Schellon S, for the ODIXa-DVT Study Investigators. Treatment of proximal deep vein thrombosis with the oral direct factor Xa inhibitor rivaroxaban (BAY 59-7939): the ODIXa-DVT (Oral Direct Factor Xa Inhibitor BAY 59-7939 in Patients With Acute Symptomatic Deep-Vein Thrombosis) study. *Circulation* 2007; **116**: 180–7.

18 Spyropoulos AC. Brave new world: the current and future use of novel anticoagulants. *Thromb Res* 2008; **123**: 29–35.

19 Bauer KA. New anticoagulants: Anti IIa vs. Anti Xa- is one better?. *J Thromb Thrombolysis* 2006; **21**: 67–72.

20 Executive Steering Committee on behalf of the SPORTIF III Investigators. Stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation (SPORTIF III): randomized controlled trial. *Lancet* 2003; **362**: 1691–8.

21 Albers GW, Diener HC, Frison L, Grind M, Nevinson M, Partridge S, Halperin JL, Horrow J, Olsson SB, Petersen P, Vahanian A. Ximelagatran vs warfarin for stroke prevention in patients with nonvalvular atrial fibrillation: a randomized trial. *JAMA* 2005; **293**: 690–8.

22 http://www.boehringer-ingelheim.com/news/news_releases/press_releases/2010/20_October_2010_dabigatran.html. Accessed 22 October 2010.

23 http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm?fuseaction=Search.DrugDetails. Accessed 22 October 2010.

24 Ufer M. Comparative efficacy and safety of the novel oral anticoagulants dabigatran, rivaroxaban and apixaban in preclinical and clinical development. *Thromb Haemost* 2010; **103**: 572–85.

25 http://www.pradaxa.com. Accessed 22 October 2010.

26 Blech S, Ebner T, Ludwig-Schwellinger E, Stangier J, Rother W. The metabolism and disposition of the oral direct thrombin inhibitor, dabigatran, in humans. *Drug Metab Dispos* 2008; **36**: 386–99.

27 Kubitza D, Becka M, Wensing G, Voith B, Zuehlsdorf M. Safety, pharmacodynamics, and pharmacokinetics of BAY 59-7939- an oral, direct Factor Xa inhibitor- after multiple dosing in healthy male subjects. *Eur J Clin Pharmacol* 2005; **61**: 873–80.
28 Weitz JI. Emerging anticoagulants for the treatment of venous thromboembolism. *Thromb Haemost* 2006; **96**: 274–84.
29 Stangier J, Rathgen K, Stahl H, Gansser D, Roth W. The pharmacokinetics, pharmacodynamics and tolerability of dabigatran etexilate, a new oral direct thrombin inhibitor, in healthy male subjects. *Br J Clin Pharmacol* 2007; **64**: 292–303.
30 Shantsila E, Lip GY. Apixaban, an oral, direct inhibitor of activated Factor Xa. *Curr Opin Invest Drugs* 2008; **9**: 1020–33.
31 Stangier J. Clinical pharmacokinetics and pharmacodynamics of the oral direct thrombin inhibitor dabigatran etexilate. *Clin Pharmacokinet* 2008; **47**: 285–95.
32 Weinz C, Schwarz T, Pleiss U, Schmeer K, Kubitza D, Mueck W, Condol D, Atken H. Metabolism and distribution of [14C] BAY 59-7939- an oral, direct Factor Xa inhibitor. *Drug Metab Rev* 2004; **36**(Suppl. 1): 98 (abstract 196).
33 RE-MODEL Study Group: Eriksson BI, Dahl OE, Rosencher N, Kurth AA, van Dijk CN, Frostick SP, Kalebo P, Christiansen AV, Hantel S, Hettiarachchi R, Schnee J, Buller HR. Oral dabigatran etexilate vs. subcutaneous enoxaparin for the prevention of venous thromboembolism after total knee replacement: the RE-MODEL randomized trial. *J Thromb Haemost* 2007; **5**: 2178–85.
34 RE-NOVATE Study Group, Eriksson BI, Dahl OE, Rosencher N, Kurth AA, Niek van Dijk C, Frostick SP, Prins MH, Hettiarachchi R, Hantel S, Schnee J, Buller HR. Dabigatran etexilate versus enoxaparin for prevention of venous thromboembolism after total hip replacement: a randomized, double-blind, non-inferiority trial. *Lancet* 2007; **370**: 949–56.
35 RE-MOBILIZE Writing Committee, Ginsberg JS, Davidson BL, Comp PC, Francis CW, Friedman RJ, Huo MH, Lieberman JR, Muntz JE, Raskob GE, Clements ML, Hantel S, Schnee JM, Caprini JA. Oral thrombin inhibitor Dabigatran etexilate vs. North American enoxaparin regimen for prevention of venous thromboembolism after knee arthroplasty surgery. *J Arthroplasty* 2009; **24**: 1–9.
36 Schulman S, Kearon C, Kakkar AK, Mismetti P, Schellong S, Eriksson H, Baanstra D, Schnee J, Goldhaber SZ, RE-COVER Study Group. Dabigatran vs. Warfarin in the Treatment of Acute Venous Thromboembolism. *N Engl J Med 2009* 2010; **361**: 2342–52.
37 Ezekowitz M, Reilly P, Nehmiz G, Simmers TA, Nagarakanti R, Parcham-Azad K, Pedersen KE, Lionetti DA, Stangier J, Wallentin L. Dabigatran with or without concomitant ASA compared with warfarin along in patients with nonvalvular atrial fibrillation (PETRO study). *Am J Cardiol* 2007; **100**: 1419–26.
38 Connolly SJ, Ezekowitz MD, DPhil CB, Yusuf S, Eikelboom J, Oldgren J, Parekh A, Pogue J, Reilly PA, Themeles E, Varrone J, Wang S, Alings M, Xavier D, Zhu J, Diaz R, Lewis BS, Darius H, Diener HC, Joyner CD, *et al*. Dabigatran vs. Warfarin in patients with Atrial Fibrillation. *N Engl J Med* 2009; **361**: 1139–51.
39 Ezekowitz MD, DPhil M, Connolly S, Parekh A, Reilly PA, Varrone J, Wang S, Oldgren J, Themeles E, Wallentin L, Tusuf S. Rationale and design of RE-LY: randomized evaluation of long-term anticoagulant therapy, warfarin, compared with Dabigatran. *Am Heart J* 2009; **157**: 805–10.
40 Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radfor MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA* 2001; **285**: 2864–70.
41 Diener HC, Connolly SJ, Ezekowitz MD, Wallentin L, Reilly PA, Yang S, Xavier D, Di Pasquale G, Yusuf S, RE-LY study group. Dabigatran compared with warfarin in patients with atrial fibrillation and previous transient ischemic attack or stroke: a subgroup analysis of the RE-LY trial. *Lancet Neurol* 2010; **9**: 1157–63.
42 http://www.clinicaltrials.gov. Accessed 9 August 2010.
43 Olsson SB, Rasmussen LH, Tveit A, Jesen E, Wessman P, Panifilov S, Wahlander K. Safety and tolerability of an immediate-release formulation of the oral direct thrombin inhibitor AZD0837 in the prevention of stroke and systemic embolism in patients with atrial fibrillation. *Thromb Haemost* 2010; **103**: 604–12.
44 Lip GYH, Rasmussen LH, Olsson SB, Jensen EC, Persson AL, Eriksson U, Wahlander KF. Oral direct thrombin inhibitor AZD0837 for the prevention of stroke and systemic embolism in patients with non-valvular atrial fibrillation: a randomized dose-guiding, safety and tolerability study of four doses of AZD0837 vs. vitamin K antagonists. *Eur Heart J* 2009; **30**: 2897–907.
45 RECORD1 Study Group, Eriksson BI, Borris LC, Friedman RJ, Haas S, Huisman MV, Kakkar AK, Bandel TJ, Beckmann H, Muehlhofer E, Misselwitz F, Geerts W. Rivaroxaban versus enoxaparin for thromboprophylaxis after hip arthroplasty. *N Engl J Med* 2008; **358**: 2765–75.
46 RECORD2 Investigators, Kakkar AK, Brenner B, Dahl OE, Eriksson BI, Mouret P, Muntz J, Soglian AG, Pap AF, Misselwitz F, Haas S. Extended duration rivaroxaban versus short-term enoxaparin for the prevention of venous thromboembolism after total hip arthroplasty: a double-blind, randomized controlled trial. *Lancet* 2008; **372**: 31–9.
47 RECORD3 Investigators, Lassen MR, Ageno W, Borris LC, Lieberman JR, Rosencher N, Bandel TJ, Misselwitz F, Turpie AG. Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty. *N Engl J Med* 2008; **358**: 2776–86.
48 Turpie AGG, Lassen MR, Davidson BL, Bauer KA, Gent M, Kwong LM, Cushner FD, Lotke PA, Berkowitz SD, Bandel TJ, Benson A, Misselwitz F, Fisher WD. Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty (RECORD 4): a randomized trial. *Lancet* 2009; **373**: 1673–80.
49 Turpie AGG, Lassen MR, Kakkar AK, Eriksson B, Misselwitz F, Bandel TJ, Homering M, Westermeier T, Gent M. A pooled analysis of four pivotal studies of rivaroxaban for the prevention of venous thromboembolism after orthopedic surgery: effect on symptomatic venous thromboembolism, death, and bleeding. *Blood* (ASH Annual Meeting Abstracts) 2008; **112**: 18–9 abstract 36.
50 Executive Steering Committee on behalf of the ROCKET AF Study Investigators. Rivaroxaban-Once daily, oral, direct factor Xa inhibition compared with vitamin K antagonism for prevention of stroke and embolism trial in atrial fibrillation: rationale and design of the ROCKET AF study. *Am Heart J* 2010; **159**: 340–7.
51 http://sciencenews.myamericanheart.org/pdfs/ROCKET_AF_pslides.pdf; Accessed 23 February 2011.
52 Lassen M, Gallus A, Raskob G, Gallus A, Pineo G, Chen D, Portman RJ. Apixaban or enoxaparin for thromboprophylaxis after knee replacement. *N Engl J Med* 2009; **361**: 594–604.
53 Lassen MR, Raskob GE, Gallus AS, Pineo GF, Chen B, Hornick P. Apixaban vs. enoxaparin for thromboprophylaxis after knee replacement (ADVANCE-2): a randomised double-blind trial. *Lancet* 2010; **9717**(375): 807–15.
54 Lopes RD, Alexander JH, Al-Khatib SM, Ansell J, Diaz R, Easton JD, Gersh BJ, Granfer CB, Hanna M, Horowitz J, Hylek EM, McMurray JJ, Verheugh FW, Wallentin L, ARISTOTLE Investigators. Apixaban for Reduction In Stroke and Other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) trial: design and rationale. *Am Heart J* 2010; **159**: 131–9.
55 http://www.bms.com/news/press_releases/. Accessed 10 June 2010.
56 http://www.escardio.org/congresses/esc-2010/congress-reports/Pages/708-3-AVERROES.aspx. Accessed 22 October 2010.
57 Fuji T, Fujita S, Tachibana S, Kawai Y. Randomized, double blind, multi-dose efficacy, safety and biomarker study of the oral factor Xa inhibitor DU-176b compared with placebo for prevention of venous thromboembolism in patients after total knee arthroplasty. *Blood (ASH Annual Meeting Abstracts)*. 2008; **112**: 34.
58 Raskob G, Cohen A, Eriksson B, McDonald A, Puskas D, Shi M, Verho M, Weitz J. Randomized, double blind, multi-dose trial of the

oral factor Xa inhibitor DU-176b versus LMW heparin (dalteparin) for the prevention of venous thromboembolism after total hip replacement. *European Heart J* 2008; **29** (Suppl. 1) Abstract 3712: 609.
59 Weitz Jl, Connolly SJ, Kunitada S, Jin J, Patel I. Randomized, parallel group, multicentre, multinational study evaluating safety of DU-176b compared with warfarin in subjects with non-valvular atrial fibrillation. *Blood (ASH Annual Meeting Abstracts)*. 2008; **112**: 33.
60 http://www.portola.com/pdfs/ESC_Betrixaban_Poster.pdf. Accessed 22 October 2010.
61 Turpie AGG. New oral anticoagulants in atrial fibrillation. *Eur Heart J* 2007; **29**: 155–65.
62 Khoo CW, Tay KH, Shantsila E, Lip G. Novel oral anticoagulants. *Int J Clin Pract, April* 2009; **63**: 630–41.
63 http://www.portola.com/PRT064445-Fxa-Inhibitor-Antidote. Accessed 22 October 2010.
64 http://www.portola.com/pdfs/ESC_Betrixaban_Poster.pdf. Accessed 22 October 2010.
65 http://www.aryx.com/wt/page/at15923. Accessed 22 October 2010.
66 Ellis DJ, Usman MH, Milner PG, Canafax DM, Ezekowitz M. The First Evaluation of a Novel Vitamin K Antagonist, Tecarfarin (ATI-5923), in Patients with Atrial Fibrillation. *Circulation* 2009; **120**: 1029–35.
67 Hylek EM, Evans-Molina C, Shea C, Henault LE, Regan S. Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. *Circulation* 2007; **115**: 268–2696.
68 The Amadeus Investigators. Comparison of idraparinux with vitamin K antagonists for prevention of thromboembolism in patients with atrial fibrillation: a randomized, open-label, non-inferiority trial. *Lancet* 2008; **371**: 315–21.
69 Healey JS, Hart RG, Pogue J, Pfeffer MA, Hohnloser SH, De Caterina R, Flaker G, Yusuf S, Connolly SJ. Risks and benefits of oral anticoagulation compared with clopidogrel plus aspirin in patients with atrial fibrillation according to stroke risk: the atrial fibrillation clopidogrel trial with irbesartan for prevention of vascular events (ACTIVE-W). *Stroke* 2008; **39**: 1482–6.
70 Camm JA, Kirchhof P, Lip GYH, Schotten U, Savelieva I, Ernst S, van Gelder IC, Al-Attar N, Hindricks G, Prendergast B, Heidbuchel H, Alfieri O, Angelini A, Atar D, Colonna P, De Caterina R, De Sutter J, Goette A, Gorenek B, Heldal M *et al*. Guidelines for the Management of Atrial Fibrillation. *Eur Heart J* 2010; **31**: 2369–429.
71 van Ryn J, Ruehl d, Priepke H, Hauel N, Wienen W. Reversibility of the anticoagulant effect of high doses of the direct thrombin inhibitor dabigatran, by recombinant factor VIIA or activated prothrombin complex concentrate. *Haematologica* 2008; **93** (Suppl. 1): 148 abstr 0370.
72 Gruber A, Marzec UM, Buetehorn U, Hanson S, Perzborn E. Potential of activated prothrombin complex concentrate and activated factor VII to reverse the anticoagulant effects of rivaroxaban in primates. *Blood* 2008; **112**: 1307. (abstr 3825).
73 Lu G, DeGuzman FR, Lakhotia S, Hollenbach SJ, Phillips DR, Sinha U. Recombinant antidote for reversal of anticoagulation by factor Xa inhibitors. *Blood* 2008; **112**: 362 (abstr 983).
74 Phillips KW, Ansell JA. The clinical implications of new oral anticoagulants: will the potential advantages be achieved? *Thromb Haemost* 2010; **103**: 34–9.
75 Yasuda K, Lan LB, Sanglard D, Furuya K, Schuetz JD, Schuetz EG. Interaction of Cytochrome P450 3A Inhibitors with P-Glycoprotein. *J Pharm Exp Ther* 2002; **303**: 323–32.
76 Lin JH, Yamazaki M. Role of P-glycoprotein in pharmacokinetics: clinical implications. *Clin Pharmacokinet* 2003; **42**: 59–98.
77 Walenga JM, Adiguzel C. Drug and dietary interactions of the new and emerging oral anticoagulants. *Int J Clin Pract* 2010; **64**: 956–67.
78 Ruff CT, Giugliano RP, Antman EM, Crugnale SE, Bocanegra T, Mercuri M, Hanyok J, Patel I, Shi M, Salazar D, McCabe CH, Braunwald E. Evaluation of the novel factor Xa inhibitor edoxaban compared with warfarin in patients with atrial fibrillation: design and rationale for the effective aNticoaGulation with factor xa next generation in atrial fibrillation–thrombolysis in myocardial infarction study 48 (ENGAGE AF–TIMI 48). *Am Heart J* 2010; **160**: 635–41. e2.
79 http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/DrugInteractionsLabeling/ucm081177.htm.
80 DuBuske LM. The Role of P-glycoprotein and organic anion-transporting polypeptides in drug interactions. *Drug Saf* 2005; **28**: 789–801.
81 Kim RB. Drugs as P-glycoprotein substrates, inhibitors and inducers. *Drug Metab Rev* 2002; **34**: 47–54.

Medical Care Research and Review
68(4) 490–503

DOI: 10.1177/1077558710396378
http://mcr.sagepub.com
SAGE

Excess Costs Attributable to Postoperative Complications

Kathleen Carey[1,5], Theodore Stefos[2,5], Shibei Zhao[1], Ann M. Borzecki[1,3,5], and Amy K. Rosen[4,5]

Abstract

This article estimates excess costs associated with postoperative complications among inpatients treated in Veterans Health Administration (VA) hospitals. The authors conducted an observational study on 43,822 hospitalizations involving inpatient surgery in one of 104 VA hospitals during fiscal year 2007. Hospitalization-level cost regression analyses were performed to estimate the excess cost of each of 18 unique postoperative complications. The authors used generalized linear modeling techniques to account for the heavily skewed cost distribution. Costs were measured using an activity-based cost accounting system and complications were assessed based on medical chart review conducted by the VA 'National Surgical Quality Improvement Program. The authors found excess costs associated with postoperative complications ranging from $8,338 for "superficial surgical site infection" to $29,595 for "failure to wean within 24 hours in the presence of respiratory complications." The results obtained suggest that quality improvement efforts aimed at reducing postoperative complications can contribute significantly to lowering of hospital costs.

Keywords

hospitals, quality, patient safety, cost

This article, submitted to *Medical Care Research and Review* on August 25, 2010, was revised and accepted for publication on December 9, 2010.

[1]VA Center for Health Quality, Outcomes and Economic Research
[2]VA Office of Productivity, Efficiency and Staffing
[3]Boston University School of Medicine
[4]VA Center for Organization, Leadership and Management Research
[5]Boston University School of Public Health

Corresponding Author:
Kathleen Carey, Boston University School of Public Health, 715 Albany Street, Boston, MA 02118, USA
Email: kcarey@bu.edu

In the decade since publication of the Institute of Medicine's landmark report "To Err is Human" (Kohn, Corrigan, & Donaldson, 2000), the quality and safety of hospital inpatient care in the United States has come under increased public scrutiny. Despite widespread interest, considerable effort expended, and a degree of measured progress, the pace of quality improvement has fallen short of the Institute of Medicine recommendations, and patient safety recently has been cited as an area in urgent need of attention (Agency for Healthcare Research and Quality [AHRQ], 2010). Mounting pressure to improve the quality and safety of hospital inpatient care is complicated by recent high-cost growth in hospital services. In the current environment of economic constraints, efficiency in both quality improvement and health care delivery is becoming an increasingly important goal.

Concern over adverse events and complications in U.S. hospitals has also stimulated a growing body of research. Some studies have used medical chart abstraction to demonstrate the extent of such events (Bates et al., 1995; Brennan et al., 1991; Classen, Pestonik, Evans, & Burke, 1991; Thomas et al., 2000), and a few have addressed their cost burden, which has been found to be substantial (Bates et al., 1997; Dimick et al., 2004; Dimick, Pronovost, Cowan, & Lipsett, 2003; Peng, Kurtz, & Johannes, 2006; Shannon et al., 2006; Thomas et al., 1999). For example, recent studies of Pennsylvania hospitals found that excess costs because of central line–associated bloodstream infections were $26,839 (Shannon et al., 2006), and for patients with hospital-acquired infections overall, charges were four times higher than for patients without hospital-acquired infections (Peng et al., 2006).

Some broader based studies of medical error costs have drawn on administrative data; a number of these have measured adverse events using the patient safety indicators (PSIs) developed by the AHRQ in response to the need for more standardized measures of patient safety. These studies also have demonstrated the high cost of safety lapses. For example, the average difference in 90-day expenditures between patients who had one of 14 PSIs and those who did not was $36,000 among 22,477 Medicare surgeries in 2000 (Encinosa & Hellinger, 2005), and ranged from $646 to $28,000 for 161,004 commercially insured surgeries in 2001-2002 (Encinosa & Hellinger, 2008). A study of the general patient population in 994 U.S. hospitals in 2000 found average excess charges of $9,000 to $58,000 for patients experiencing a PSI (Zhan & Miller, 2003), and an analysis of 2002 Medicare patient data estimated the extra payment for five adverse events measured using PSIs at $300 million per year (Zhan, Friedman, Mosso, & Pronovost, 2006). A recent large-scale study using administrative data on a 40-state sample from the period 1998 to 2006 developed coding rules for identifying hospital-acquired infections, and found excess costs in the range of $33,000 to $46,000 for hospitalizations involving surgery (Eber, Laxminararyan, Perencevich, & Malani, 2010). Finally, a study of PSI adverse events in Veterans Health Administration (VA) hospitals in 2001 found excess costs in the range of $5,000 to $40,000 per hospitalization (Rivard et al., 2008).

New Contribution

While all postoperative complications are not the result of medical error, complications are considered valid indicators of the general quality of inpatient care

(Lawthers et al., 2000; Weingart et al., 2000). They have been found to be among the most frequent of inpatient adverse events, as well as the most costly (Thomas et al., 1999), since complications following surgery can lead to increased days of hospitalization, greater intensity of services, more ancillary services, and/or extra medications. We sought to estimate the excess costs of postoperative complications in surgical inpatients treated in VA hospitals during fiscal year 2007. Our study is unique in that it simultaneously (a) assesses the occurrence of complications through measures obtained via chart abstraction rather than through imperfect algorithms applied to administrative data, (b) relies on hospitalization-level cost estimates as opposed to charge or reimbursement data, (c) examines a broad set of postoperative complications occurring in a large number of hospitals, and (d) compares excess costs obtained using administrative data with excess costs obtained using medical record data. In this study, we conceptualize costs as the resources that hospitals expend to produce their services. While other research perspectives may be valuable, we maintain that under current U.S. health care reform it is essential to achieve a significant slowing of cost growth at the point of health care delivery and that a focus on the hospital's prospects for cost control is highly pertinent to the ongoing policy discussion.

Method

This is an observational study on a sample of 47,040 hospitalizations among patients who underwent inpatient surgery in one of the 104 acute-care VA hospitals during fiscal year 2007 (October 1, 2006, to September 30, 2007). The unit of observation is the unique hospitalization. Patient-level cost regressions are performed on binary variables indicating the presence of a surgical complication, controlling for severity, demographics, and facility-level variables.

Data and Variables

The VA has two distinct systems for determining the cost of a hospitalization. The VA Health Economics Resource Center takes a top-down approach that allocates the VA national budget using coefficients from a Medicare cost function linking cost-adjusted charges to patient demographic and hospitalization information (Chapko et al., 2009; Wagner, Chen, & Barnett, 2003). In contrast, the bottom-up Decision Support System (DSS) is an activity-based costing system that creates relative value units at the local departmental level to determine the cost of intermediate products that are in turn summed to the level of the individual hospitalization (Barnett, 2003; Chapko et al., 2009), independent of patient mix, market conditions, or socioeconomic characteristics. While both systems capture costs from the hospital's perspective at the level of the hospitalization, the greater patient specificity accounted for in the DSS approach is more likely to be better suited to gauging high patient costs because of specific complications experienced by unique patients than the Health Economics Resource Center costs (which are averaged over patients sharing the same characteristics). For

this reason, we used the DSS costs in this study. The dependent variable was the log of the DSS measured cost of the hospitalization.

The key independent variable was a binary indicator of whether the patient experienced one of 19 postoperative complications. Data were obtained from the VA National Surgical Quality Improvement Program (VASQIP), which VA commenced in 1991 in response to Congressional mandate for VA to compare its risk-adjusted surgical outcomes with those of private hospitals. Program founders believed that postoperative outcomes were particularly relevant for measuring general quality of surgical care, since a surgical operation is a predictable event with an expected outcome in most cases. While they were not particularly focused on preventability, they believed that good science promotes good management, and targeted 30-day adverse outcomes as a means toward quality improvement (Henderson & Daley, 2009). The VASQIP uses a comprehensive medical record review to collect data that include 19 distinct postoperative complications. The VASQIP assesses a systematic random sample of VA major noncardiac surgeries, accounting for approximately 70% of VA major surgeries, and 25% of VA surgeries overall (Henderson & Daley, 2009). The complication variables are generic in nature, because the program was intended to cover a broad scope of surgical specialties. Medical record data are considered the "gold standard" for measuring adverse events and complications, because they are obtained directly from patients' medical charts (Best et al., 2002).

We followed previous literature for estimating the excess cost associated with PSIs in VA by controlling for severity, demographics, and facility-level variables commonly associated with variation in patient cost (Rivard et al., 2008). This required that we merge the VASQIP data with the VA Patient Treatment File (PTF), an inpatient administrative file containing information on all VA hospitalizations. Because the VASQIP data were organized at the surgical procedure level and the PTF data were organized at the hospitalization level, we merged only those records in which VASQIP surgery dates fell between PTF admission and discharge dates.

We used the Elixhauser comorbidity index, constructed using PTF data, as a proxy for illness severity (Elixhauser, Steiner, Harris, & Coffey, 1998). This index was constructed in two steps. First, for all hospitalizations in the database, we estimated a fixed effects regression of the natural logarithm of costs on age and binary variable indicators for 27 comorbid conditions. The fixed effects represented unique clinical classifications measured by diagnosis-related groups. Second, we summed the regression coefficients for the comorbid conditions that were significant at the 5% level ($p < .05$) for each hospitalization according to whether or not the comorbid condition was present during the hospitalization.

Demographic control variables in the cost regressions included four age categories and gender, obtained from the PTF. Facility-level controls were indicator variables constructed from the American Hospital Association 2007 Annual Survey Database for major teaching facilities and indicator variables for facilities located in metropolitan areas. Major teaching status was defined by membership in the Council of Teaching Hospitals. Metropolitan areas were defined using the Census Bureau's Core Based

Statistical Areas as urban areas with populations of 50,000 or more. Finally, we included hospital level and diagnosis-related group fixed effects.

Analytic Technique

The assumptions underlying estimators generated by ordinary least squares regression are generally not met in analyses of individual hospitalization costs, which usually exhibit significant positive skewness and heteroscedasticity, violations likely amplified in cases characterized by postoperative complications. Researchers often address this problem with log transformation of costs. However, raw dollars is normally the scale of interest, and retransformation is not straightforward, since the expected value of the log of dependent variable y conditional on independent variable x is not likely to be equal to the log of the expected value of y conditional on x. Suitable retransformation techniques have been developed (Duan, Manning, Morris, & Newhouse, 1983) but significant bias in inference can arise if heteroscedasticity is not characterized and applied to the retransformation process (Manning, 1998; Mullahy, 1998). We took an alternative approach to estimation by using generalized linear modeling (GLM), which provides appropriate estimators of mean values when the distribution of data is nonnormal. GLM models have the practical advantage of providing estimates of the conditional mean directly, without requirement of retransformation (Basu & Manning, 2009).

We estimated GLMs corresponding to each of the unique complications, using a gamma-distributed log-linked structure, selected on the basis of the Modified Park Test (Manning & Mullahy, 2001). For each model, we evaluated the excess cost because of the complication by first assuming that every hospitalization had the complication associated with it and predicting the cost associated with that hospitalization. We then assumed that no hospitalization in the sample had the complication and predicted the cost. The excess cost associated with the complication was calculated as the difference between the predicted cost with the complication and the predicted cost without the complication (Encinosa & Hellinger, 2005, 2008; Rivard et al., 2008; Zhan & Miller, 2003). We estimated models using PROC GENMOD, SAS Version 9.1.

Results

Of the 47,040 observations (unique hospitalizations), we omitted 1,930 observations for which the patient experienced more than one complication. This exclusion permitted us to avoid confounding estimates of excess cost across the range of specific complications studied. This left 45,110 observations, of which 44,911 had complete data. From these, we excluded 610 observations on hospitalizations for which the patient was transferred to another hospital. Of the remaining 44,301 hospitalizations, 479 resulted in death. Table 1 presents descriptive statistics on the sample of 44,301 observations, which included patients who died in hospital and on the sample of 43,822 observations on hospitalizations for patients who were alive at discharge.

Table 1. Descriptive Statistics[a]

Variable	Mean or Proportion (Standard Deviation)	
	All Hospitalizations	Alive at Discharge
Dependent		
Cost of hospitalization (dollars)	25,461 (30,760)	25,865 (26,592)
Independent		
Age		
18 years and younger (%)	3.19	3.22
8-39 years (%)	55.32	55.53
40-64 years (%)	21.74	21.72
65 years or older (%)	19.75	19.52
Female gender (%)	5.14	5.18
Comorbidity index	0.09 (0.14)	0.08 (0.13)
	N = 44,301	*N* = 43,822

a. Excludes transfer patients.

For analysis purposes, we excluded the complication "coma lasting longer than 24 hours," because only four hospitalizations had this complication and the patient died in all four cases. Table 2 presents the remaining 18 postoperative complications, their VASQIP definitions, the percentage of hospitalizations in which the complication was present and for which the patient was transferred to another hospital or died, and the frequencies of events. Frequencies are listed for the two samples used in the analyses: all hospitalizations on nontransferred patients and hospitalizations on nontransferred patients discharged alive. The events are relatively infrequent, in the latter group ranging from 0.34 per 1,000 for "cardiac arrest requiring CPR" (cardiopulmonary resuscitation) to 8.28 per 1,000 for "superficial surgical site infection."

Model results for the 18 complication indicators associated with excess costs are presented in Table 3. For the sample of all hospitalizations, the coefficients of the complication variables are positive and highly significant, with two exceptions: "peripheral nerve injury" and "cardiac arrest requiring CPR." The former complication exhibits no statistical significance, likely because of low frequency of occurrence. The latter complication is negatively and significantly associated with cost. As the majority of patients in this group (77.4%) died, this result suggests that death occurred early in the hospitalization, such that relatively low costs were incurred. Among the 16 complications that were significantly related to cost, the calculated excess costs ranged from $8,234 for "progressive renal insufficiency" to $28,779 for "failure to wean from ventilator within 48 hours." The directions and associations of the covariates (not shown) were as expected. Cost rose with age and increased level of the comorbidity index.

The final two columns of Table 3 list the coefficients and excess costs for hospitalizations in which the patient was discharged alive. The results are generally similar to

Table 2. Postoperative Complications: Descriptions and Frequencies

VASQIP Variable Label	Percentage of Hospitalizations (With Complications) That Ended in Transfer or Death		Mean Number of Events per Thousand Surgical Patient Hospitalizations Excluding Transfers	
	Transfer	Death	All Hospitalizations[a]	Alive at Discharge[b]
Central nervous system complications				
Cardiac arrest requiring CPR	1.41	77.4	2.45	0.34
Myocardial infarction	8.72	25.6	1.85	1.48
Cerebral vascular accident/stroke	9.76	26.8	1.27	1.05
Peripheral nerve injury	1.96	0.00	0.96	0.98
Other surgical complications				
Bleeding requiring >4 units PRBCs	5.21	22.8	2.12	1.83
Deep vein thrombosis/ thrombophlebitis	4.23	17.5	1.51	1.44
Graft/prosthesis failure	3.42	7.69	1.56	1.53
Systemic sepsis	4.83	31.5	4.79	4.11
Respiratory complications				
Failure to wean >48 hours	4.14	30.2	5.34	4.13
Pneumonia—outcome	2.66	22.2	7.90	7.46
Pulmonary embolism	2.31	16.2	1.27	1.23
Reintubation: respiratory/cardiac failure	5.10	32.5	4.16	3.65
Urinary tract complications				
Acute renal failure (postoperative)	0.00	21.9	0.71	0.57
Progressive renal insufficiency	5.83	30.0	2.32	1.99
Urinary tract infection	3.77	11.6	8.46	8.28
Wound complications				
Wound disruption/dehiscence	3.20	12.1	2.36	2.35
Superficial surgical site infection	2.42	4.28	8.17	8.24
Deep wound surgical site infection	2.04	12.7	2.49	2.53

Note: CPR = cardiopulmonary resuscitation; VASQIP = VA National Surgical Quality Improvement Program; PRBC = packed red blood cell.
a. N = 44,301.
b. N = 43,822.

those of the larger sample that includes patients who died, although excess costs are somewhat higher, reflecting longer periods of hospitalization during which excess costs were incurred. For "pulmonary embolism" and "reintubation for respiratory/cardiac

Table 3. GLM Results for Log Costs and Estimated Excess Cost Per Hospitalization

	All Hospitalizations		Alive at Discharge	
VASQIP Variable Label	Coefficient (Standard Error of Coefficient)	Excess Cost Estimate[a]	Coefficient (Standard Error of Coefficient)	Excess Cost Estimate[a]
Central nervous system complications				
Cardiac arrest requiring CPR	–0.1693 (0.0572**)	–3,975	0.0286 (0.1523)	722
Myocardial infarction	0.2636 (0.0698)**	7,688	0.3300 (0.0738)**	9,729
Cerebral vascular accident/stroke	0.6193 (0.0834)**	21,852	0.6305 (0.0879)**	21,858
Peripheral nerve injury	0.1308 (0.0910)	3,564	0.1309 (0.0901)	3,485
Other surgical complications				
Bleeding requiring > 4 units PRBCs	0.2926 (0.0630)**	8,661	0.3349 (0.0622)**	9,897
Deep vein thrombosis/ thrombophlebitis	0.4352 (0.0746)**	13,891	0.4444 (0.0746)**	13,918
Graft/prosthesis failure	0.4273 (0.0727)**	13,582	0.4303 (0.0726)**	13,375
Systemic sepsis	0.6476 (0.0420)**	23,087	0.6658 (0.0442)**	23,422
Respiratory complications				
Failure to wean from ventilator within 48 hours	0.7616 (0.0445)**	28,779	0.7880 (0.0445)**	29,595
Pneumonia—outcome	0.3931 (0.0322)**	12,222	0.4164 (0.0329)**	12,798
Pulmonary embolism	0.4883 (0.0796)**	16,052	0.4309 (0.0803)**	13,411
Reintubation for respiratory/cardiac failure	0.4513 (0.0451)**	14,501	0.3969 (0.0469)**	12,104
Urinary tract complications				
Acute renal failure (postoperative)	0.5704 (0.1054)**	19,598	0.6003 (0.1181)**	20,474
Progressive renal insufficiency	0.2800 (0.0613)**	8,234	0.2936 (0.0635)**	8,491
Urinary tract infection	0.4222 (0.0313)**	13,321	0.4364 (0.0312)**	13,542
Wound complications				
Wound disruption/ dehiscence	0.5027 (0.0587)**	16,628	0.5137 (0.0584)**	16,688
Superficial surgical site infection	0.2896 (0.0318)**	8,537	0.2897 (0.0316)**	8,338
Deep wound surgical site infection	0.5408 (0.0568)**	18,256	0.5458 (0.0563)**	18,036

Note: GLM = generalized linear modeling; CPR = cardiopulmonary resuscitation; VASQIP = VA National Surgical Quality Improvement Program; PRBC = packed red blood cell.
a.Mean value in dollars.
**$p < .001$.

Table 4. Excess Costs: VASQIP and PSI Estimates Compared

VASQIP Measures		PSI Measures[a]		
Variable	Excess Cost Estimate: Regression Method ($)	Variable	Excess Cost Estimate: Regression Method ($)	Excess Cost Estimate: Case Matching Method ($)[b]
Failure to wean	29,595	Respiratory failure	12,360	50,953
Reintubation	12,104			
Pulmonary embolism	13,411	Pulmonary embolism/ deep vein thrombosis	11,620	9,237
Deep vein thrombosis	13,918			
Sepsis	23,422	Sepsis	17,172	40,080
Wound dehiscence	16,688	Wound dehiscence	22,154	24,236

Note: PSI = patient safety indicator; VASQIP = VA National Surgical Quality Improvement Program.
a. From Rivard et al. (2008), inflated to 2007 dollars using the Bureau of Labor Statistics producer price index for hospital services.
b. Case matching approach was included in Rivard et al. (2008) in addition to the regression approach to replicate the methods of Zhan and Miller (2003).

failure," however, excess costs were higher among patients who died, suggesting some unmeasured degree of severity for these complications. Finally, patients with "cardiac arrest requiring CPR" who were discharged alive did not incur significant excess costs.

In Table 4, we compare our estimates with those of aforementioned earlier research on excess cost in VA that used PSIs to measure adverse events (Rivard et al., 2008). Table 4 lists excess cost estimates of those postoperative PSI events that correspond to VASQIP complications among patients discharged alive. In 2007 dollars, our excess cost estimate for "failure to wean" was 139% higher than the regression-estimated PSI excess cost for postoperative respiratory failure but within 2% of the regression-estimated PSI excess cost for postoperative respiratory failure in the case of "reintubation." For "pulmonary embolism" and "deep vein thrombosis," our estimates were approximately 18% higher than the comparable PSI regression-estimated excess cost, and for "sepsis," approximately 36% higher. Our results for "wound dehiscence," on the other hand, were substantially lower (approximately 25%) than the PSI regression-estimated excess cost.

Discussion

Assessing the costs associated with adverse patient events is complicated by difficulties in measuring the incidence of these events, as well as the lack of standardized data for a comprehensive national evaluation of patient safety (AHRQ, 2010). Moreover, the development of valid and feasible measures of the cost of health care events has

proven to be very difficult (Lipscomb, Yabroff, Brown, Lawrence, & Barnett, 2009). This article addresses both of these challenges by drawing on high-quality, standardized measures of the incidence and cost of inpatient postoperative complications collected nationally by the VA.

Our analysis contributes new information to understanding comparative costs, which varied considerably over a broad range of complications. We found excess costs associated with postoperative complications in VA to be considerable, ranging from 3% to 120% higher (for "cardiac arrest requiring CPR" and "failure to wean," respectively) than inpatient costs in the absence of complications among patients discharged alive. The low cost associated with cardiac arrests suggests that survivors are likely to be relatively healthy respondents who have a transient postoperative arrhythmia. While studies of PSIs have generally found sepsis more expensive than postoperative respiratory complications at least in part because of longer hospital stays, it is possible that the higher costs seen among this sample in patients with "failure to wean" was because they were sicker going into the procedure, that is, more likely to be undergoing emergency surgery, with less time to optimize preexisting conditions, than those suffering some of the other complications.

Comparison of VASQIP events with those of corresponding postoperative PSIs revealed considerable discrepancies in both magnitude and direction of estimated excess costs. Hence, our results do not provide evidence that using measures derived from algorithms applied to administrative data are good substitutes for those using the more labor-intensive chart-abstracted complication rates when estimating excess cost. At the same time, it should be borne in mind that the VASQIP and PSI definitions are not precisely comparable, and this may account for some of the discrepancy between the two measures. PSIs are defined using ICD-9 coding, whereas VASQIP complications are defined by clinical definitions applied by nurse abstractors based on reviews of physician documentation as well as laboratory and radiologic data (Romano et al., 2009). Moreover, because of lack of Present on Admission (POA) coding in VA administrative data, our PSI calculations do not account for POA, which may explain some of the differences, particularly with "deep vein thrombosis" and "pulmonary embolism" (Houchens, Elixhauser, & Romano, 2008). VASQIP measures, on the other hand, provide a gold standard for examination of the issue, and accordingly, can distinguish between complications that arose following surgery and complications that were present on admission. Finally, the previously cited Rivard et al. (2008) study using PSIs differed from the present study in that it used Health Economics Resource Center rather than DSS costs. Nevertheless, further study using PSIs would be worthwhile, because even though identification of adverse events using administrative data has some limitations, efficiency in their construction (yielding millions of observations compared with thousands for the more expensive approach) may outweigh the lower reliability of screening for adverse events. Moreover, with POA data elements available, the validity of PSIs constructed using administrative data is improved.

Our estimates are informative in prioritizing efforts to improve quality of care in VA hospitals and also in determining whether financial incentives are well aligned versus misaligned with improvement. Results suggest that directing efforts toward

reducing complications such as cerebral vascular accidents, systemic sepsis, acute renal failure, and failure to wean, each of which incurred excess costs of greater than $20,000, might have high value. However, that postoperative complications have high costs must be interpreted in light of the fact that not all complications reflect deficiencies in care. High complication rates can be reflective of underlying severity of disease in the patient population (Silber & Rosenbaum, 1997). There is also some evidence that hospitals known for quality may have higher rates of adverse events because they pay more attention to documenting this (Isaac & Jha, 2008). Moreover, the cost per event is not in itself sufficient to render it high priority, as some complications bring about greater patient harm than others. Quality improvement efforts must consider the cost-effectiveness of reducing particular complications. Such evaluation requires estimates of the frequency and morbidity associated with these events, the total attributable cost, and also the cost of the prevention activities themselves. Many complications cannot be avoided, and among those that can, prevention may not be cost-effective for events that are very rare or exceedingly costly to prevent. Nevertheless, the results of our study are informative in demonstrating the range of potential returns to investment in quality improvement in VA hospitals.

There are limitations to this study. The excess costs that we measured were only those incurred during the inpatient stay. However, the excess costs extend beyond the initial hospitalization, so that any interventions undertaken will appear to be more cost-effective if postoperative costs following discharge are taken into account (Encinosa & Hellinger, 2008). From a methodological perspective, it should be borne in mind that this is an observational study using data from a single year. No method can completely identify causality from observational data. Without a randomized controlled experiment, conclusions regarding the effect of variables of interest on outcomes are subject to parameter estimate bias to the extent of correlation between measured variables of interest and omitted variables that are significantly associated with outcomes. Finally, our study was of VA hospitals only, using VA-specific cost measures, so that excess costs observed here do not necessarily generalize to nonfederal U.S. hospitals.

As U.S. hospitals respond to concern over patient safety by investing in quality improvement efforts, knowledge of the relative cost-saving potential of reducing various inpatient complications becomes increasingly important. Future research should continue to refine the measurement of patient safety and cost as well as improve understanding of the relative potential for prevention across the range of adverse events.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

This material is based upon work supported by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Health Services Research and

Development (IIR 07-072), Kathleen Carey Principal Investigator. The authors would like to acknowledge the VA Surgical Quality Data Use Group (SQDUG) for its role as scientific advisors and for the critical review of data use and analysis presented in this manuscript. The opinions expressed are those of the authors and not necessarily those of the Department of Veterans Affairs or the United States Government.

References

Agency for Healthcare Research and Quality. (2010, March). *National Health Care Quality Report 2009* (AHRQ Publication No. 10-0003). Retrieved from http://www.ahrq.gov/qual/nhqr09/nhqr09.pdf

Barnett, P. G. (2003). Determination of VA health care costs. *Medical Care Research and Review*, *60*(3, Suppl.), 124S-141S.

Basu, A., & Manning, W. G. (2009). Issues for the next generation of health care cost analyses. *Medical Care*, *47*(7, Suppl. 1), S109-S114.

Bates, D. W., Cullen, D. J., Laird, N., Petersen, L. A., Small, S. D., Servi, D., . . . the ADE Prevention Study Group. (1995). Incidence of adverse drug events and potential adverse drug events. *Journal of the American Medical Association*, *274*, 29-34.

Bates, D. W., Spell, N., Cullen, D. J., Burdick, E., Laird, N., Petersen, L. A., . . . Leape, L. L. (1997). The cost of adverse drug events in hospitalized patients. *Journal of the American Medical Association*, *277*, 307-311.

Best, W. R., Khuri, S. F., Phelan, M., Hur, K., Henderson, W. G., Demakis, J. G., & Daley, J. (2002). Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: Results from the Department of Veterans Affairs National Surgical Quality Improvement Program. *Journal of the American College of Surgeons*, *194*, 257-266.

Brennan, T. A., Hebert, L. E., Laird, N. M., Lawthers, A., Thorpe, K. E., Leape, L. L., . . . Hiatt, H. H. (1991). Hospital characteristics associated with adverse events and substandard care. *Journal of the American Medical Association*, *265*, 3265-3269.

Chapko, M. K., Liu, C. F., Perkins, M., Li, Y. F., Fortney, J. C., & Maciejewski, M. L. (2009). Equivalence of two healthcare costing methods: Bottom-up and top-down. *Health Economics*, *18*, 1188-1201.

Classen, D. C., Pestonik, S. L., Evans, R. S., & Burke, J. P. (1991). Computer surveillance of adverse drug events in hospital patients. *Journal of the American Medical Association*, *266*, 2847-2851.

Dimick, J. B., Chen, S. L., Taheri, P. A., Henderson, W. G., Khuri, S. F., & Campbell, D. A. (2004). Hospital costs associated with surgical complications: A report from the private-sector National Surgical Quality Improvement Program. *Journal of the American College of Surgeons*, *199*, 531-537.

Dimick, J. B., Pronovost, P. J., Cowan, J. A., & Lipsett, P. A. (2003). Complications and costs after high-risk surgery: Where should we focus quality improvement initiatives? *Journal of the American College of Surgeons*, *196*, 671-678.

Duan, N., Manning, W. G., Morris, C. N., & Newhouse, J. P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics*, *1*, 115-126.

Eber, M. R., Laxminarayan, R., Perencevich, E. N., & Malani, A. (2010). Clinical and economic outcomes attributable to health care-associated sepsis and pneumonia. *Archives of Internal Medicine*, *170*, 347-353.

Elixhauser, A., Steiner, C., Harris, R., & Coffey, R. M. (1998). Comorbidity measures for use with administrative data. *Medical Care*, *36*, 8-27.

Encinosa, W. E., & Hellinger, F. J. (2005). What happens after a patient safety event? Medical expenditures and outcomes in Medicare. In K. Henriksen, J. B. Battles, E. S. Marks, & D. I. Lewin (Eds.), *Advances in patient safety: From research to implementation: Vol. 1. Research findings* (pp. 423-446). Rockville, MD: Agency for Healthcare Research and Quality.

Encinosa, W. E., & Hellinger, F. J. (2008). The impact of medical errors on ninety-day costs and outcomes: An examination of surgical patients. *Health Services Research*, *43*, 2067-2085.

Henderson, W. G., & Daley, J. (2009). Design and statistical methodology of the National Surgical Quality Improvement Program: Why is it what it is? *American Journal of Surgery*, *198*(5, Suppl.), S19-S27.

Houchens, R. L., Elixhauser, A., & Romano, P. S. (2008). How often are potential patient safety events present on admission? *Joint Commission Journal on Quality and Patient Safety*, *34*, 154-163.

Isaac, T., & Jha, A. K. (2008). Are patient safety indicators related to widely used measures of hospital quality? *Journal of General Internal Medicine*, *23*, 1373-1378.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (2000). *To err is human: Building a safer health system* (A report of the Committee on Quality of Health Care in America, Institute of Medicine). Washington, DC: National Academies Press.

Lawthers, A. G., McCarthy, E. P., Davis, R. B., Peterson, L. E., Palmer, R. H., & Iezzoni, L. I. (2000). Identification of in-hospital complications from claims data: Is it valid? *Medical Care*, *38*, 785-795.

Lipscomb, J., Yabroff, K. R., Brown, M. L., Lawrence, W., & Barnett, P. G. (2009). Health care costing: Data, methods, current applications. *Medical Care*, *47*(7, Suppl. 1), S1-S6.

Manning, W. G. (1998). The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics*, *17*, 283-295.

Manning, W. G., & Mullahy, J. (2001). Estimating log models: To transform or not to transform? *Journal of Health Economics*, *20*, 461-494.

Mullahy, J. (1998). Much ado about two: Reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics*, *17*, 247-281.

Peng, M. M., Kurtz, S., & Johannes, R. J. (2006). Adverse outcomes from hospital-acquired infection in Pennsylvania cannot be attributed to increased risk on admission. *American Journal of Medical Quality*, *21*(6, Suppl.), 17S-28S.

Rivard, P. E., Luther, S. L., Christiansen, C. L., Zhao, S., Loveland, S., Elixhauser, A., . . . Rosen, A. K. (2008). Using patient safety indicators to estimate the impact of potential adverse events on outcomes. *Medical Care Research and Review*, *65*, 67-87.

Romano, P. S., Mull, H. J., Rivard, P. E., Zhao, S., Henderson, W. G., Loveland, S., . . . Rosen, A. K. (2009). Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. *Health Services Research*, *44*, 182-204.

Shannon, R. P., Patel, B., Cummins, D., Shannon, A. H., Ganguli, G., & Lu, Y. (2006). Economics of central line-associated bloodstream infections. *American Journal of Medical Quality*, *21*(6, Suppl.), 7S-16S.

Silber, J. H., & Rosenbaum, P. R. (1997). A spurious correlation between hospital mortality and complication rates. *Medical Care*, *35*(10, Suppl.), OS77-OS92.

Thomas, E. J., Studdert, D. M., Burstin, H. R., Orav, E. J., Zeena, T., Williams, E. J., . . . Brennan, T. A. (2000). Incidence and types of adverse events and negligent care in Utah and Colorado. *Medical Care*, *38*, 261-271.

Thomas, E. J., Studdert, D. M., Newhouse, J. P., Zbar, B. I. W., Howard, K. M., Williams, E. J., & Brennan, T. A. (1999). Costs of medical injuries in Utah and Colorado. *Inquiry*, *36*, 255-264.

Wagner, T. H., Chen, S., & Barnett, P. G. (2003). Using average cost methods to estimate encounter-level costs for medical-surgical stays in the VA. *Medical Care Research and Review*, *60*(3, Suppl.), 15S-36S.

Weingart, S. N., Iezzoni, L. I., Davis, R. B., Palmer, R. H., Cahalane, M., Hamel, M. B., . . . Banks, N. J. (2000). Use of administrative data to find substandard care: Validation of the complications screening program. *Medical Care*, *38*, 796-806.

Zhan, C., Friedman, B., Mosso, A., & Pronovost, P. (2006). Medicare payment for selected adverse events under the Prospective Payment System: Building the business case for investing in patient safety improvement. *Health Affairs*, *25*, 1386-1393.

Zhan, C., & Miller, M. R. (2003). Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. *Journal of the American Medical Association*, *290*, 1868-1874.

Positive Predictive Value of the AHRQ Patient Safety Indicator "Postoperative Sepsis": Implications for Practice and Policy

Marisa Cevasco, MD, MPH, Ann M Borzecki, MD, MPH, Qi Chen, MPH, Patricia A Zrelak, PhD, CNRN, CNAA-BC, Marlena Shin, JD, MPH, Patrick S Romano, MD, MPH, Kamal MF Itani, MD, FACS, Amy K Rosen, PhD

BACKGROUND: Patient Safety Indicator (PSI) 13, or "Postoperative Sepsis," of the Agency for Healthcare Quality and Research (AHRQ), was recently adopted as part of a composite measure of patient safety by the Centers for Medicare and Medicaid Services (CMS). We sought to examine its positive predictive value (PPV) by determining how well it identifies true cases of postoperative sepsis.

STUDY DESIGN: Two retrospective cross-sectional studies of hospitalization records that met PSI 13 criteria were conducted, one within the Veterans Administration (VA) Hospitals from fiscal years (FY) 2003 to 2007, and one within community hospitals between October 1, 2005 and March 31, 2007. Trained abstractors reviewed medical records from each database using standardized abstraction instruments. We determined the PPV of the indicator and performed descriptive analyses of cases.

RESULTS: Of 112 cases flagged and reviewed within the VA system, 59 were true events of postoperative sepsis, yielding a PPV of 53% (95% CI 42% to 64%). Within the community hospital sector, of 164 flagged and reviewed cases, 67 were true cases of postoperative sepsis, yielding a PPV of 41% (95% CI 28% to 54%). False positives were due to infections that were present on admission, urgent or emergent cases, no clinical diagnosis of sepsis, or other coding limitations such as nonspecific shock in postoperative patients.

CONCLUSIONS: PSI 13 has relatively poor predictive ability to identify true cases of postoperative sepsis in both the VA and nonfederal sectors. The lack of information on diagnosis timing, confusion about the definition of elective admission, and coding limitations were the major reasons for false positives. As it currently stands, the use of PSI 13 as a stand-alone measure for hospital reporting appears premature. (J Am Coll Surg 2011;212:954–961. © 2011 by the American College of Surgeons)

In 2003, the Agency for Healthcare Research and Quality (AHRQ) developed a formal set of Patient Safety Indicators (PSIs) to screen administrative databases for gaps in patient safety. Specifically, the PSIs are ICD-9-CM-based algorithms that use hospital discharge data to screen for safety-related inpatient events. Initially, the PSIs were designed for case-finding and quality improvement activities. However, the PSIs are increasingly being used to assess hospital performance and as part of pay-for-performance programs.[1,2] Nine states use PSIs for public reporting on hospitals; the National Quality Forum formally endorsed 10 PSIs for hospital performance measures, and the Centers for Medicare and Medicaid Services have adopted 4 PSIs and 1 composite measure to compare quality and safety across hospitals.[3,4]

Despite the increasing use of PSIs in assessing hospital performance, validation studies investigating the positive predictive value (PPV) of these indicators demonstrate wide variability in their accuracy to detect true events com-

Disclosure Information: Nothing to disclose.

This work was supported in part by the US Agency for Healthcare Research and Quality contract 290-04-0020.

Received August 27, 2010; Revised November 27, 2010; Accepted November 29, 2010.

From the Department of Surgery, VA Boston Healthcare System (Cevasco, Itani); the Department of Surgery, Brigham and Women's Hospital (Cevasco); the Department of Health Policy and Management, Boston University School of Public Health (Borzecki, Chen, Rosen); Boston University School of Medicine (Borzecki, Itani); the Center for Organization, Leadership, and Management Research, VA Boston Healthcare System (Chen, Shin, Rosen); and Harvard Medical School (Itani), Boston, MA; the Center for Health Quality, Outcomes and Economic Research, Bedford VAMC, Bedford, MA (Borzecki); and the Center for Healthcare Policy and Research (Zrelak) and the Division of General Medicine and Center for Healthcare Policy and Research (Romano);University of California, Davis, CA.

Correspondence address: Marisa Cevasco, MD, MPH, Department of Surgery, Brigham and Women's Hospital, 75 Francis St, Boston, MA 02115.

ISSN 1072-7515/11/$36.00
doi:10.1016/j.jamcollsurg.2010.11.013

Abbreviations and Acronyms

AHRQ = Agency for Healthcare Research and Quality
IRR = inter-rater reliability
PPV = positive predictive value
PSI = Patient Safety Indicator
VA = Veterans Health Administration

pared with the "gold standard" of medical record review. PPVs ranging between 44% and 91% have been documented for various PSIs.[5-11] This variation has been attributed to ambiguity in coding guidelines, differences in coding practices between hospitals, and the inability of ICD-9 CM codes to distinguish between diagnoses that were present on admission versus those that developed after admission.[6]

In this study, we focused on PSI 13, postoperative sepsis, and determined its PPV in both the Veterans Health Administration (VA) and the nonfederal sector. Rates of postoperative sepsis will continue to be investigated because they are part of the AHRQ PSI composite measure used by the Centers for Medicare and Medicaid Services to compare quality and safety across hospitals. It is a significant complication, associated with a 21.9% excess mortality, 10.2 extra days in the hospital, and nearly $60,000 in excess charges.[12] Many cases of postoperative sepsis may be prevented through appropriate use of perioperative antibiotics, good surgical site preparation, and sterile surgical techniques.[13]

This is the first study to investigate the PPV of the postoperative sepsis PSI and the first study to directly compare the PPV of a single PSI in 2 different health care systems. In this article, we examine the PPV of PSI 13, characterize the type of events incorrectly captured by this PSI, and assess which perioperative events may contribute to postoperative sepsis. Results from this study will have important implications for hospital reporting, pay-for-performance, and other uses of this specific PSI.

METHODS

PSI definition

PSI 13, postoperative sepsis, captures all discharges of patients age 18 or older who underwent an elective operating room procedure with a postoperative length of stay greater than 3 days and at least 1 of 21 ICD-9-CM diagnosis codes with sepsis as a secondary diagnosis (Table 1). Excluded from this definition are patients admitted with a principal diagnosis of infection or sepsis, a secondary diagnosis of infection or sepsis POA, patients admitted for a nonelective hospitalization (used as a proxy for urgent or emergent surgical intervention), patients with cancer or immunocompromised states, and all obstetric admissions (pregnancy, childbirth).

Table 1. ICD-9-CM Sepsis Diagnosis Codes Present in Any Secondary Diagnosis Field

ICD-9-CM sepsis diagnosis code	Definition
0380	Streptococcal septicemia
0381	Staphylococcal septicemia
03810	Staphylococcal septicemia
03811	Methicillin-susceptible *Staphylococcus aureas* septicemia
03812	Methicillin-resistant *Staphylococcus aureas* septicemia
03819	Other staphylococcal septicemia
0382	Pneumococcal septicemia
0383	Septicemia due to anerobes
78552	Septic shock
78559	Shock without mention of trauma
9980	Postoperative shock
99591	Systemic inflammatory response syndrome due to infectious process without organ dysfunction
99592	Systemic inflammatory response syndrome due to infectious process with organ dysfunction
Septicemia due to	
03840	Gram-negative organism, unspecified
03841	Hemophilus influenza
03842	*Escherichia coli*
03843	Pseudomonas
03844	Serratia
03849	Septicemia due to other gram-negative organisms
0388	Other unspecified septicemia
0389	Unspecified septicemia

VA methods

For the VA analysis, we applied AHRQ PSI software (v.3.1.a) to VA fiscal years 2003 to 2007 administrative data to identify patients with suspected postoperative sepsis. The required Institutional Review Board approvals from the Bedford VA Medical Center and the VA Boston Healthcare System were obtained to conduct this study.

Hospital selection

We randomly selected 28 of 150 acute care VA hospitals to obtain a manageable number of hospitals for individual medical record review while capturing variation in coding across facilities. Our sampling scheme was designed to generate a sample of hospitals with a range of observed and expected safety events. The scheme of randomization and selection has been simultaneously published by our group.[11] Twenty-one percent of sampled

hospitals were in the Northeast, 32% were in the South, 18% were in the Midwest, and 29% were in the West. Eighty-nine percent were major teaching hospitals (resident-to-bed ratio $>$ 0.25). The median number of hospital beds was 155 (range 62 to 360 beds).

Case selection

We randomly selected 4 medical records flagged with postoperative sepsis from each of the 28 hospitals, for a total of 112 medical records. Based on previously reported PPV estimates, 112 cases per PSI were selected to ensure reasonably narrow PPV confidence intervals ($\pm$ 10% to 20%).

Medical record abstraction and inter-rater reliability

To determine rates of true and false positives for the PSI, 2 trained nurses conducted a retrospective chart review using standardized chart abstraction tools and guidelines developed by AHRQ. These were modified for the VA's electronic medical record and to achieve inter-rater reliability (IRR) measurement standards of greater than 90% observed agreement. Specifically, medical records were reviewed for occurrence of postoperative sepsis; demographic characteristics, comorbidities, and risk factors of the sampled patients; and patient outcomes after the event, such as transfer to a higher level of care, development of end organ dysfunction, or death. The presumed source(s) of infection leading to sepsis, the presence or absence of positive blood cultures, and the procedures performed during the hospital stay were also abstracted from the medical record. Perioperative processes were noted, including administration of preoperative antibiotics and postoperative glycemic control. Nurse training included several sessions discussing the rationale of PSI 13, the likely sources of information needed from the electronic medical record, and a systematic chronology for chart abstraction. We also conducted IRR testing to ensure standardized and reliable abstraction. Before independent abstraction, a series of medical records were reviewed by both nurses, and IRR was measured as the percentage of agreement on 40 selected key clinical questions. IRR testing ultimately revealed 95% agreement after 3 rounds of 5 records each.

Community hospital methods

The study was approved by the federal Office of Management and Budget and by the Institutional Review Board at the University of California, Davis Medical Center (UCDMC). Each participating hospital was provided with a Notice of Data Use, indicating that all collected data would be considered confidential unless otherwise compelled by law. As in the VA analysis, AHRQ PSI software version 3.1.a was used.

Hospital selection

Hospitals were recruited for participation in the study through the AHRQ Quality Indicators LIST-SERV. Forty-seven hospitals representing 29 states agreed to participate on a voluntary basis and without financial compensation. They included not-for-profit, nonreligious hospitals (78%), hospitals with a religious affiliation (4%), and for-profit hospitals (4%). Large hospitals with more than 400 beds constituted more than one-third of participants (36%), small hospitals with fewer than 200 beds constituted 21% of participants, and no participant had fewer than 50 beds.

Case selection

Discharges between October 1, 2005 and March 31, 2007 were used to extract a sample of medical records that met PSI 13 criteria. The number of postoperative sepsis cases varied among hospitals, and the sampling fraction at each hospital was adjusted to achieve a target sample size of 240 cases (recognizing that some records would be unavailable or ineligible for abstraction).

Medical record abstraction

A data abstraction instrument was developed as noted above. Study staff pretested the abstraction instrument at UC Davis Medical Center and 2 nonprofit hospitals in the Sacramento region. Additional input was obtained from national experts as needed. The entire instrument and corresponding guidelines are available online at http://qualityindicators.ahrq.gov/validationpilot.htm.

Analysis

Positive predictive value

We calculated PPV as the rate of true positives divided by the total number of medical records reviewed and derived 95% confidence intervals about that estimate.

True positive analysis

For patients with confirmed postoperative sepsis, we performed descriptive analyses of multiple continuous and categorical variables including demographic characteristics (age, sex, race or ethnicity), comorbidities, relevant risk factors, and the nature of the surgical procedure and outcomes. All statistical analyses were performed using SAS version 9.1 (SAS Institute, Inc.).

False positive analysis

All false positive cases underwent further detailed review to better understand why they were incorrectly flagged by the PSI algorithms. However, the authors did not have access to the original records, so cases that were reclassified as "nonelective" were designated as false positive even if the record was correctly coded.

Table 2. Characteristics of True Positive Patients

Characteristic	VA sample (n = 59)	Community hospital sample (n = 67)
Demographics and comorbid conditions		
Male gender, n (%)	57 (97)	48 (78)
Age, y, mean	64	65
Current cigarette smoker, n (%)	19 (32)	19 (29)
Diabetes, n (%)	22 (37)	14 (21)
Type of case, n (%)		
Cardiothoracic	12 (20)	12 (18)
Vascular	11 (19)	6 (9)
General and gastrointestinal surgery	11 (19)	12 (18)
Documentation of preoperative antibiotic use, n (%)	48 (81)	67 (100)
Use of razor, n (%)	20 (34)	17 (25)
Source of infection, n (%)		
Pneumonia	35 (59)	36 (54)
Urinary tract infection	17 (29)	8 (12)
Catheter-related bloodstream infection	13 (22)	31 (46)
Surgical site infection	9 (15)	1 (2)

VA, Veterans Health Administration

RESULTS

VA hospitals

Within the 28 selected hospitals, 33,548 cases met the eligibility criteria for postoperative sepsis during the study period, and 218 cases were flagged by PSI 13. Of the 112 reviewed cases, 59 were confirmed as postoperative sepsis, yielding a PPV of 53% (95% CI 42% to 64%).

True positive analysis

The sample was 97% male (n = 57), with a mean age of 63.8 years. Seventy-six percent of patients were white, 9% were African American, and 7% were Hispanic. The average body mass index was 30.0 kg/m^2. Thirty-seven percent of patients had diabetes and 32% were smokers. There was no IV drug abuse documented among these patients, no patient had a chronic indwelling IV catheter in place on admission, and 2 patients used steroids within the month before admission.

The majority of cases were cardiothoracic (n = 12), vascular (n = 11), or general surgery (n = 11). Thirty-three cases (56%) were labeled as clean, 16 cases (27%) were clean-contaminated, and 1 case was contaminated (the remainder were missing case classification information). Preoperative antibiotics were given in 48 cases (81%). Eleven patients (19%) either did not receive antibiotics or documentation was absent. If hair removal was necessary, a razor was used in 20 instances (34%) and clippers in 15 (25%).

Average blood glucose level at 6:00 AM on postoperative day 1 was 154 mg/dL. A single patient with blood glucose over 200 mg/dL had no antiglycemic ordered; otherwise patients were appropriately treated with oral hypoglycemic agents or intravenous or subcutaneous insulin.

Regarding the source of infection, 59% of cases were attributed to pneumonia, 29% to a urinary tract infection, 22% to a catheter-related bloodstream infection, and 15% to a surgical site infection. These diagnoses were not mutually exclusive. Blood cultures were positive in 26 patients (44%), but 6 patients (10%) did not have blood cultures sent despite having a documented diagnosis of sepsis (Table 2).

Patient outcomes

The average length of stay was 32 days. Slightly more than half the patients required a transfusion of packed red cells or plasma (n = 31), and 16 (27%) patients required transfer to a higher level of care due to postoperative sepsis. All-cause in-hospital mortality was 35% (n = 21) (Table 3).

False positive analysis

Of 53 false positive cases, 16 (30%) patients had an infection present on admission and 12 (23%) were emergent admissions. Seven (13%) cases of sepsis were diagnosed after admission but before the index procedure. Twelve patients (23%) had postoperative sepsis listed as a possible diagnosis in the medical record but no antibiotics were ever administered, no imaging was notable for an infection, nor were any blood cultures sent. Other false positives represented limitations of the ICD-9-CM coding system in that some patients who had postoperative hemorrhagic shock (n = 3) or cardiogenic shock (n = 3) were assigned a diagnosis code of 998.0 (postoperative shock) in the absence of a more specific code to describe their condition (Table 4).

Community hospitals

From the 49 selected hospitals, we identified 164 cases of postoperative sepsis. Of these, 67 cases were true events of

Table 3. Patient Outcomes

Consequences of postoperative sepsis	VA sample	Community hospital sample
Positive blood cultures, n (%)	44 (26)	30 (45)
Move to higher level of care, n (%)	27 (16)	Not abstracted
Blood product transfusion, n (%)	53 (31)	33 (22)
Length of stay, d, mean	32	27
All-cause mortality, n (%)	21 (35)	19 (28)

VA, Veterans Health Administration

Table 4. Characteristics of False Positives

Characteristic	VA sample (n = 53)		Community hospital sample (n = 97)	
	n	%	n	%
Infection present on admission	16	30	27	28
Nonelective admission	12	23	41	43
Not confirmed by medical record	12	23	27	28
Diagnosis after admission but before index procedure	7	13	Not evaluated	
Other coding-related inaccuracies	6	11	0	
Incompleteness of the medical record	0		2	2

VA, Veterans Health Administration

postoperative sepsis, yielding a PPV of 41% (95% CI 28% to 54%).

True positive analysis

The sample was 48% male (n = 32), with a mean age of 65 years. Twenty-one percent of patients had diabetes and 29% were smokers. Four patients had used oral steroids within the 60 days before admission. One patient had chronic liver disease and 10 patients had chronic pulmonary disease.

The majority of cases were cardiothoracic (n = 12, 18%), general surgery (n = 12, 18%), or vascular surgery (n = 6, 9%). Preoperative antibiotics were given in 64 patients (96%) within 60 minutes of incision time. The remaining 3 patients had antibiotics given within 15 minutes of the 1-hour standard. If hair removal was necessary, the use of a razor was noted in 25% of cases.

Average blood glucose at 6:00 AM on postoperative day 1 was 111 mg/dL. Regarding source of infection, 54% were attributed to pneumonia, 46% to a catheter-related bloodstream infection, 12% to a urinary tract infection, and 2% to a surgical site infection. These diagnoses were not mutually exclusive (Table 2).

Patient outcomes

The average length of stay was 27 days. Greater than 92% of patients developed some evidence of systemic infection or end organ dysfunction, including a temperature greater than 38° C (92% of patients); systolic blood pressure ≤ 95 mmHg (92% of patients); postoperative blood glucose greater than 180 mg/dL (61% of patients); and low urine output (50% of patients). Twenty-two patients received a blood product transfusion. All-cause in-hospital mortality was 28% (n = 19). Deaths attributed to sepsis based on clinical documentation in the medical record represented 14% (n = 9) of patients (Table 3).

False positive analysis

Of 97 false positive cases, 43% were nonelective admissions, 28% had infection or sepsis present on admission, and 28% had no documentation of bacteremia, septicemia, sepsis, or systemic inflammatory response syndrome. The majority of patients admitted with infection or sepsis were readmissions from an earlier procedure or complications from a previous inpatient stay (Table 4).

DISCUSSION

One of the most important goals of developing the PSIs was to accurately identify potentially preventable complications of medical care. This goal, representing the criterion validity of the indicator, can be assessed by determining the PPV of each PSI. We found that the PPV of PSI 13 was low in both the VA and nonfederal sectors: specifically, 53% (95% CI 42% to 64%) in the VA sector and 41% (95% CI 28% to 54%) in the nonfederal sector. Recent studies have demonstrated PPVs between 44% and 91% for other PSIs.[5-11] Our PPV estimates for PSI 13 are at the low end of this range. The similarity of these estimates, despite the lack of financial incentives in the VA system, suggests that factors inherent to coding, such as ambiguous codes, play an important role in the accuracy of this PSI based on administrative databases.[6]

Investigation of the false positive cases confirmed that the low criterion validity of PSI 13 stems primarily from diagnosis timing and coding issues. In the VA sample, nearly one-third of false positives were cases of sepsis that were present on admission. Similarly in the nonfederal sector, 28% of false positives were present on admission. These findings are consistent with previous studies that showed that introducing a "present on admission flag" (as many state health data agencies and the Centers for Medicare and Medicaid Services have done) could substantially decrease the high false positive rate for several PSIs.[6] By eliminating cases that had sepsis present on admission, the PPV of the indicator would improve to 67% and 56% in the VA and nonfederal sectors, respectively. However, these values are arguably still poor. So, implementation of a present on admission flag alone is insufficient to improve the criterion validity of this indicator.

Other reporting and coding issues contributed to poor criterion validity. Patients with urgent or emergent operations accounted for a significant percentage of false positive cases (23% in the VA sector and 43% in the nonfederal sector), although abstractors often reported some confusion about the distinction between elective and nonelective procedures. Some patients were incorrectly coded as having

sepsis when there was no clinical diagnosis of sepsis (23% in the VA sector and 28% of the nonfederal sector). Some of these latter cases stemmed from coding the "rule out" or "versus" diagnoses of sepsis referred to in a physician's note when a postoperative patient experienced hypotension or other signs of end organ dysfunction (cardiogenic shock, hemorrhagic shock, hypovolemia, intravascular volume depletion, acute renal failure, etc). Although this coding practice is well justified, it decreases the specificity of this PSI.

Coding issues and missing information about diagnosis timing are not solely to blame for the low PPV of this indicator. One inherent difficulty with this PSI is that there are varying clinical definitions of sepsis. Providers may have different thresholds and methods of diagnosing a patient as septic. For the purpose of this study, we relied solely on physician documentation of sepsis within the medical record, regardless of whether the physician adhered to a standardized definition of sepsis. Depending on the threshold of the physician, patients with the same clinical state may not be given the same diagnosis. Moreover, a revised consensus definition of sepsis was published during the study period, which likely contributed to variation in the classification of sepsis.[14] This ambiguity may have contributed to the high number of "rule out" sepsis diagnoses that were incorrectly captured under this PSI. It is important to note that 2 organisms typically associated with community-acquired infections (*H. influenza* and *Pneumococcus sp.*) are included in the list of 21 ICD-9-CM codes that are used to capture postoperative sepsis. This highlights that some cases of postoperative sepsis captured under this indicator may be due to community-acquired organisms, making it difficult to attribute all cases of postoperative sepsis to nosocomial infections.

Another concern about PSI 13 is that it potentially describes events also captured by PSI 7 (central venous catheter-related bloodstream infections). In our study, approximately one-quarter of the VA patients and nearly half of the nonfederal patients had sepsis attributed to an infected vascular catheter. PSI 7 was developed to capture infections due to central venous catheters, but does not exclude postoperative cases of sepsis.[15] Under the current PSI classification system, cases of postoperative sepsis stemming from a central venous catheter could be assigned both PSI 7 and 13. Although this does not diminish the criterion validity of either PSI 7 or 13, hospitals have the potential to be penalized more than once for the same complication.

A final potential shortcoming of this indicator is that the prevalence of postoperative sepsis is low, making this PSI less reliable for judging hospital performance than some of the more common PSIs. However, PSI 13 does detect clinically consequential events, as demonstrated by the high rates of bacteremia, systemic inflammatory response syndrome, and mortality seen in both sets of true positive patients. This meets an important goal of the AHRQ Patient Safety Indicators, which is to identify patients for whom there is a potential opportunity to improve clinical outcomes.[16] In the VA sample, nearly 50% of patients were bacteremic, 27% required transfer to a higher level of care such as an ICU, and all-cause mortality was 35%. Similarly, for the nonfederal sector, 92% of patients developed evidence of end organ dysfunction and all-cause mortality was 28%. These high rates of adverse outcomes are reflected in the mean length of stay—32 days for true positive cases within the VA and 27 days for true positive cases within the nonfederal sector. Moreover, a recent study within the VA system demonstrated that relative costs per inpatient hospitalization were 2.28 times greater for patients with sepsis relative to patients without sepsis. This difference in risk-adjusted costs amounted to nearly $28,000 (Vaughan-Sarrazin and colleagues. Costs of post-operative sepsis: the business case for quality improvement to reduce postoperative sepsis in VA hospitals. Personal communication, 2011).

Although we are limited by the retrospective nature of this study, analysis of the true cases of postoperative sepsis suggests that at least some infections leading to postoperative sepsis may be preventable. Nosocomial pneumonia was cited as the most frequent cause of infection in both the VA and nonfederal sectors. We know that judicious postoperative pain control, limiting the use of nasogastric tubes, good oral hygiene, preoperative optimization of COPD, keeping the head of bed at greater than 30 degrees in intubated patients, and expeditious weaning to extubate have been associated with lower rates of nosocomial pneumonia. We also know that hospital-acquired pneumonia can increase a patient's hospital stay by more than 1 week, resulting in up to a 3-fold increase in mortality.[17] Similar logic may be applied to the other causes of postoperative sepsis identified in this study, such as catheter-related bloodstream infections, urinary tract infections, and surgical site infections. With respect to surgical site infection, it is striking that razors were used in 20 operative cases in the VA sector (34%) and 17 cases in the nonfederal sector (20%), despite consensus guidelines established in 1999 dictating the use of clippers.[18] Although this study did not investigate postoperative processes of care, it does suggest the need for vigilance to prevent these postoperative infections.

Study limitations

There are some inherent limitations to our study. The patient populations were different in the VA and nonfederal sectors, making direct comparisons difficult. Similarly, the data were abstracted from 2 different but overlapping time periods.

However, these differences in populations and time may be perceived as a relative strength of the study because the poor criterion validity of the PPV remained constant across dissimilar populations and time periods.

Another limitation is that the data elements collected and abstraction methods were not identical. For instance, within the VA sample, we were unable to obtain specific information on the timing of antibiotic administration and whether antibiotics were redosed intraoperatively. These data points are available only on paper-based anesthesia records maintained at each VA hospital. On the other hand, the nonfederal data abstraction instrument was not subject to IRR testing, nor were clinicians available to review ambiguous cases. As a result, PPV was underestimated in the community hospital sample, because any case reclassified by the abstractor as a nonelective admission was automatically labeled as false positive, even if the patient was correctly coded as having postoperative sepsis. Some shortcomings were noted in both abstraction tools; for instance, neither instrument specified whether urinary tract infections were catheter-associated. This information would have been useful when considering the perioperative processes of care that contributed to postoperative sepsis.

In the VA study, 112 cases were abstracted from a total of 218 cases that qualified for PSI 13. In the nonfederal hospitals, 164 cases that qualified for PSI 13 were reviewed. The decision to limit the VA study to 112 cases was based on the ability to allocate trained nurse-reviewers for data abstraction. However, both the VA and nonfederal studies were designed to generate confidence intervals of similar width, so this difference should not detract from the findings.

Two final limitations of this study should be noted. Using the medical record as the gold standard to determine whether the PSI coding algorithm identifies true events relies on the completeness of physician documentation. Lastly, we are unable to report on sensitivity, specificity, or negative predictive value of this PSI because we did not abstract charts of patients not flagged by the PSI 13 algorithm. Future research efforts could examine these issues because this PSI may be missing clinically significant events.

In conclusion, PSI 13 has relatively poor predictive ability to identify postoperative sepsis in both the VA and community hospital sectors, although its predictive value would be substantially higher with information on diagnosis timing (as now required by the Centers for Medicare and Medicaid Services and many state health data agencies) and greater clarity about elective admissions. Despite the low PPV of this PSI, it does identify clinically significant events, representing opportunities for real quality improvement. However, as it currently stands, the use of PSI 13 on a stand-alone basis for hospital safety profiling, public reporting and pay-for-performance measures appears premature.

Author Contributions

Study conception and design: Cevasco, Borzecki, Chen, Shin, Romano, Itani, Rosen
Acquisition of data: Chen, Shin, Zrelak
Analysis and interpretation of data: Cevasco, Borzecki, Chen, Itani
Drafting of manuscript: Cevasco, Itani
Critical revision: Cevasco, Borzecki, Rosen, Itani, Romano

Acknowledgment: We would like to thank Haytham Kaafarani, MD, for his help in developing the VA abstraction tool, and Sally MacDonald, RN and Kathleen Hickson, RN for their help as trained nurse abstractors.

REFERENCES

1. Healthgrades patient safety methodology. Available at: http://www.healthgrades.com/media/DMS/pdf/PatientSafetyMethodology.pdf. Accessed July 26, 2010.
2. Agency for Healthcare Research and Quality (AHRQ). AHRQ Quality Indicators: Guide to Patient Safety Indicators. Version 3.0a. Rockville, MD. March 2003; Version 3.1 (March 12, 2007).
3. National Quality Forum. NQF endorsed standards. Available at: http://www.qualityforum.org/Measures_List.aspx. Accessed July 26, 2010.
4. CMS. Details for: quality measures for reporting in fiscal year 2009 for 2010 update. Available at: http://www.cms.hhs.gov/apps/media/press/factsheet.asp?Counter=3044&intNumPerPage=10&checkDate=&checkKey=&srchType=1&numDays=3500&srchOpt=0&srchData=&keywordType=All&chkNewsType=6&intPage=&showAll=&pYear=&year=&desc=false&cboOrder=date. Accessed July 26, 2010.
5. Gallagher B, Cen L, Hannan E. Validation of AHRQ's patient safety indicator for accidental puncture or laceration. In: Henriksen K, Battles J, Marks E, et al eds. Advances in Patient Safety: From Research to Implementation. Concepts and Methodology. Vol 2. Rockville, MD: Agency for Healthcare Research and Quality; 2005. AHRQ Publication No. 05-0021-2.
6. White RH, Sadeghi B, Tancredi DJ, et al. How valid is the ICD-9-CM based AHRQ patient safety indicator for postoperative venous thromboembolism? Med Care 2009;47:1237–1243.
7. Utter GH, Zrelak PA, Baron R, et al. Positive predictive value of the AHRQ Accidental Puncture or Laceration Patient Safety Indicator. Ann Surg 2009;250:1041–1045.
8. White RH, Garcia M, Sadeghi B, et al. Evaluation of the predictive value of ICD-9-CM coded administrative data for venous thromboembolism in the United States. Thrombosis Research 2010;126:61–67.
9. Utter GH, Cuny J, Sama P, et al. Detection of postoperative respiratory failure: How predictive is the AHRQ Patient Safety Indicator? J Am Coll Surg 2010;211:347–351.
10. Zrelak PA, Sadeghi B, Utter GH, et al. Positive predictive value of the AHRQ Patient Safety Indicator for central line associated-bloodstream infection ("Selected infections due to medical care"). J Healthcare Quality 2011;33:29–36.

11. Kaafarani HMA, Borzecki AM, Itani KMF, et al. Validity of selected patient safety indicators: opportunities and concerns. J Am Coll Surg 2011;212:924–934.

12. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA 2003;290:1868–1874.

13. Bratzler DW, Houck PM, for the Surgical Infection Prevention Guidelines Writers Workgroup. Antimicrobial prophylaxis for surgery: an advisory statement from the National Surgical Infection Prevention Project. Clin Infect Dis 2004;38:1706–1715.

14. Levy MM, Fink MP, Marshall JC, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Intensive Care Med 2003;29:530–538.

15. AHRQ quality indicators: Guide to patient safety indicators. PSI #7 Central Venous Catheter-related Bloodstream Infections. Available at: http://www.qualityindicators.ahrq.gov/psi_download.htm. Accessed July 26, 2010.

16. Rosen AK, Rivard P, Zhao S, et al. Evaluating the patient safety indicators: how well do they perform on Veterans Health administrative data? Med Care 2005;43:873–884.

17. Siegel RE. Emerging gram-negative antibiotic resistance: daunting challenges, declining sensitivities, and dire consequences. Respir Care 2008;53:471–479.

18. Mangram AJ, Horan TC, Pearson ML, et al. Guideline for prevention of surgical site infection, 1999. Infect Control Hosp Epidemiol 1999;20:247–278.

Positive Predictive Value of the AHRQ Patient Safety Indicator "Postoperative Wound Dehiscence"

Marisa Cevasco, MD, MPH, Ann M Borzecki, MD, MPH, David A McClusky III, MD, Qi Chen, MPH, Marlena H Shin, JD, MPH, Kamal MF Itani, MD, FACS, Amy K Rosen, PhD

BACKGROUND: The Agency for Healthcare Research and Quality patient safety indicator (PSI) 14, or "postoperative wound dehiscence," is 1 of 4 PSIs recently adopted by the Centers for Medicare & Medicaid Services to compare quality and safety across hospitals. We determined how well it identifies true cases of postoperative wound dehiscence by examining its positive predictive value (PPV).

STUDY DESIGN: A retrospective cross-sectional study of hospitalization records that met PSI 14 criteria was conducted within the Veterans Health Administration hospitals from fiscal years 2003 to 2007. Trained abstractors used standardized abstraction instruments to review electronic medical records. We determined the PPV of the indicator and performed descriptive analyses of cases.

RESULTS: Of the 112 reviewed cases, 97 were true events of postoperative wound dehiscence, yielding a PPV of 87% (95% CI 79% to 92%). Sixty-one percent (n = 59) of true positive cases had at least 1 risk factor, such as low albumin level, COPD, or superficial wound infection. False positives were due to coding errors, such as cases in which the patient's abdomen was intentionally left open during the index procedure.

CONCLUSIONS: PSI 14 has relatively good predictive ability to identify true cases of postoperative wound dehiscence. It has the highest PPV among all PSIs evaluated within the Veterans Health Administration system. Inaccurate coding was the reason for false positives. Providing additional training to medical coders could potentially improve the PPV of this indicator. At present, this PSI is a promising measure for both quality improvement and performance measurement; however, its use in pay-for-performance efforts seems premature. (J Am Coll Surg 2011;212: 962–967. © 2011 by the American College of Surgeons)

Postoperative wound dehiscence is a serious surgical complication. Up to 3% of laparotomy incisions are associated with dehiscence, and more than half of repaired laparotomy dehiscences will go on to form incisional hernias, entering many patients into a cycle of surgical repair, reherniation, and acute and chronic wound complications.[1] Postoperative wound dehiscence is associated with an additional 9 days of hospitalization, $40,000 in excess charges, and 10% in-hospital attributable mortality.[2] Despite the high morbidity and mortality associated with this complication, its incidence has remained relatively unchanged over time.[3]

Postoperative wound dehiscence is defined by the Agency for Healthcare Quality and Research (AHRQ) as "reclosure of postoperative disruption of abdominal wall" in a patient who has undergone an abdominopelvic operating room procedure. Postoperative wound dehiscence may be prevented through appropriate surgical technique, optimizing modifiable patient risk factors prior to elective surgery and close monitoring of perioperative conditions.[4] Surgeon experience level and technical factors have also been shown to affect the rate of wound dehiscence.[5] Patient characteristics such as age, pulmonary disease, malnourishment, steroid use, diabetes mellitus, and obesity are each independently associated with increased rates of wound dehiscence. Perioperative conditions such as wound infection, intra-abdominal infection, and hypotension have also

Disclosure information: Nothing to disclose.

Received October 4, 2010; Revised January 16, 2011; Accepted January 19, 2011.

From the VA Boston Healthcare System, Boston, MA (Cevasco, Chen, Shin, Rosen); the Department of Surgery, Brigham and Women's Hospital, Boston, MA (Cevasco); the Center for Healthcare Quality, Outcomes and Economic Research, Bedford VA Medical Center, Bedford, MA (Borzecki); the Department of Health Policy and Management, Boston University School of Public Health (Borzecki, Chen, Rosen) and School of Medicine (Borzecki, Itani), Boston, MA; the Department of Surgery, Emory University Hospital and VA Atlanta Healthcare System, Atlanta, GA (McClusky); and Harvard Medical School, Boston, MA (Itani).

Correspondence address: Marisa Cevasco, MD, MPH, Department of Surgery, VA Boston Healthcare System, West Roxbury Division, 1400 VFW Parkway, West Roxbury, MA 02132. email: mcevasco@partners.org

ISSN 1072-7515/11/$36.00
doi:10.1016/j.jamcollsurg.2011.01.053

Abbreviations and Acronyms

AHRQ = Agency for Healthcare Quality and Research
PPV = positive predictive value
PSI = patient safety indicator
VA = Veterans Affairs
VHA = Veterans Health Administration

been associated with increased rates of postoperative wound dehiscence.[6]

Wound dehiscence is a significant complication of surgical care, occurs at a clinically meaningful rate, and may be prevented by good processes of care.[7] Because of these factors, particularly its potential preventability, postoperative wound dehiscence was selected as a patient safety indicator (PSI 14) by AHRQ. The PSIs were developed in 2003 to improve methods for identifying potential gaps in patient care related to patient safety. Originally intended for use in quality improvement initiatives, the PSIs are currently applied in ways inconsistent with this purpose. At present, more than 100 organizations are using the PSIs for national, state, and regional public reporting and hospital profiling. More recently, the Centers for Medicare & Medicaid Services added 4 individual PSIs and a composite PSI measure to their Hospital Compare Web site.[8] PSI 14 was 1 of 4 individual indicators adopted. Increasing interest in these activities, including their use by the Centers for Medicare & Medicaid Services, makes it essential that the PSIs accurately reflect hospital safety performance. Research regarding their validity is urgently needed.

We therefore conducted a study to determine the positive predictive value (PPV) of PSIs within the Veterans Health Administration (VHA). Although several of the PSIs have undergone extensive validation in both the federal and private sectors, this is the first study, to our knowledge, that explicitly evaluated the validity of this PSI using medical record data as the "gold standard." We also characterized the type of events incorrectly captured by this PSI and assessed which perioperative events may have contributed to postoperative wound dehiscence. Results from this study will have important implications for hospital reporting, pay for performance, and other uses of this specific PSI nationwide.

METHODS

Study design

We conducted a retrospective observational analysis of VHA inpatient administrative data that met criteria for PSI 14. We applied the AHRQ PSI software (version 3.1a) to VHA fiscal years 2003 to 2007 (October 1, 2003 through September 30, 2007) administrative data to identify cases with suspected postoperative wound dehiscence. The required IRB approvals from the Veterans Administration (VA) Boston Healthcare System and Bedford VA Medical Center were obtained to conduct this study.

PSI 14, "postoperative wound dehiscence," captures all discharges of patients who underwent an abdominopelvic operating room procedure and subsequent "reclosure of postoperative disruption of abdominal wall" (ICD-9-CM code 54.61). Excluded from this definition are patients who underwent reclosure procedures performed on the same day as the original surgery, those with a length of stay less than 2 days, patients with immunocompromised states, and all obstetric admissions (eg, pregnancy, childbirth).[7]

Hospital selection

We randomly selected 28 of 158 acute care VA hospitals to obtain a manageable number of hospitals for individual medical record review while capturing variation in coding across facilities. Our sampling scheme was designed to generate a sample of hospitals with a range of observed and expected safety events; previous work describes the sampling methodology in greater detail.[9] Twenty-one percent of sampled hospitals were from the Northeast, 32% were from the South, 18% were from the Midwest, and 29% were from the West. Eighty-nine percent were major teaching hospitals (ie, resident-to-bed ratio >0.25). The median number of hospital beds was 155 (range 62 to 360 beds).

Case selection

We randomly selected 4 medical records flagged with postoperative wound dehiscence from each of the 28 hospitals for a total of 112 medical records. Based on previously reported PPV estimates, 112 cases per PSI were selected to ensure reasonably narrow PPV CIs ($\pm$10% to 20%).

Medical record abstraction and inter-rater reliability (IRR)

To determine rates of true and false positives for the PSI, 2 trained nurses conducted a retrospective chart review using standardized chart abstraction tools and guidelines developed by AHRQ. These were modified to be suitable for the VHA's electronic medical records and to achieve IRR measurement standards of greater than 90% observed agreement. Specifically, medical records were reviewed for the occurrence of postoperative wound dehiscence; in addition, demographic characteristics were abstracted. For true positives, nurses also abstracted information on comorbidities; risk factors of the sampled patients; and patient outcomes following the event such as number of reparative surgeries, postoperative wound infection, or death. Additional variables abstracted included the type of surgical incision; surgical technique, and type of suture used to close

the fascia; presence of infection within the abdomen or pelvis; level of training of the person performing the procedure; and whether retention sutures were placed.

Nurse training included several sessions discussing the rationale of PSI 14, the likely sources of information needed from the electronic medical record, and a systematic chronology for chart abstraction. We also conducted IRR testing to ensure standardized and reliable abstraction. Prior to independent abstraction, a series of medical records were reviewed by both nurses, and IRR was measured as the percentage of agreement on 48 clinical questions. IRR testing ultimately revealed 96% agreement after 4 rounds of 4 records each.

Analysis

PPV

We calculated PPV as the rate of true positives divided by the total number of medical records reviewed and derived 95% CIs.

True positive analysis

For patients with confirmed postoperative wound dehiscence, we performed descriptive analyses of multiple continuous and categoric variables, including demographic characteristics (age, gender, race/ethnicity), comorbidities, risk factors, nature of the surgical procedure, and outcomes. All statistical analyses were performed using SAS version 9.1 (SAS Institute Inc).

False positive analysis

All false positive cases underwent further detailed review to better understand why they were incorrectly flagged by the PSI algorithms.

RESULTS

From the 28 selected hospitals, 191 of 29,721 eligible patients were flagged as having PSI 14. This generated an observed rate of wound dehiscence of 6.4 per 1,000 discharges, which is comparable to the national VHA rate of 6.2 per 1,000 discharges at risk. Of these 191 patients, 112 were chosen for detailed review (see case selection outlined earlier). Ninety-seven cases were true positives, yielding a PPV of 87% (95% CI 79% to 92%).

True positive analysis

The sample was 99% male (n = 96), with a mean age of 67.2 ± 11 years. Seventy-two percent were white, 7% were African American, and 7% were Hispanic. Twenty-eight percent of patients (n = 27) had COPD, 14% (n = 14) were obese (ie, body mass index >30), and 9% (n = 9) used systemic steroids. Twenty-three percent of patients (n = 23) had a preoperative albumin level of <3 g/dL (Table 1).

Table 1. True Positive Characteristics

	True positives (n = 97)
Demographics and comorbid conditions	
Male sex, n (%)	96 (99)
Mean age, y	67 ± 11
COPD, n (%)	27 (28)
Body mass index >30, n (%)	14 (14)
Systemic steroids, n (%)	9 (9)
Albumin <3 g/dL, n (%)	23 (23)
Procedure characteristics, n (%)	
Sigmoidectomy	11 (11)
Small bowel resection	6 (6)
Right hemicolectomy	6 (6)
Other	74 (77)*
Wound dehiscence characteristics, n (%)	
Evisceration	41 (43)
Fluid drainage	27 (28)
Wound infection	9 (9)
Coughing	12 (13)
Visceral leak	5 (5)
Postoperative course, n (%)	
1 additional surgery	82 (84)
≥2 surgeries	13 (13)
Median length of stay, d	26
All-cause mortality, n (%)	11 (11)

*Other procedures included left hemicolectomy, total proctocolectomy, low anterior resection, abdominoperineal resection, Hartman takedown procedure, colostomy revision, abdominal aortic aneurysm repair, right hepatectomy, paraesophageal hernia repair, radical cystectomy with ileal conduit, and open appendectomy.

Characteristics related to the index procedure

Of the 55 separate ICD-9-CM procedure codes listed as the index procedure, the most frequent were codes for sigmoidectomy (n = 11 [11%]), partial small bowel resection (n = 6 [6%]), and right hemicolectomy (n = 6 [6%]) (Table 1). Overall, one-third of the index cases involved interventions upon the small bowel, colon, and/or rectum. Twenty-five percent of index procedures were considered emergent or urgent.

During the index procedure, the majority of incisions were midline vertical (95%). The most common method of abdominal wound closure involved primary continuous closure (n = 70 [72%]) using a large polydiaxone suture (n = 67 [69%]). Interrupted techniques using absorbable (n = 12 [12%]) and nonabsorbable (n = 5 [5%]) sutures were also noted. Internal retention sutures of the Smead-Jones technique were placed in 3% (n = 3) of index pro-

cedures, and true retention sutures were placed in 1% (n = 1) of index procedures. All 4 of these procedures with retention suture placement involved a perforated hollow viscus in an obese patient who presented with fluid leaking from the wound. For patients for which this information was recorded, an attending physician performed closure in 31% (n = 26) of cases and a physician-in-training performed the closure in 69% (n = 56).

Characteristics related to the wound dehiscence

Sixty-three percent (n = 61) of wound dehiscences were diagnosed within 7 days after the index procedure. Clinically evident evisceration of abdominal contents was noted in 41 patients (43%), and 28% of patients (n = 27) had large amounts of fluid draining from the wound. Excessive coughing and/or physical exertion was noted in 13% of patients (n = 12), and a visceral leak was found in 5% of patients (n = 5). A wound infection was found in 9% of patients (n = 9) (Table 1). Of those 9 patients, 3 had a perforated hollow viscus, but the skin had been closed with either staples (n = 2) or running monocryl suture (n = 1). Wound dehiscence was attributed to fascial tearing in 32% of patients, necrotic fascia in 12%, breakage of suture material in 11%, intra-abdominal infection in 9%, and unraveling of a tied suture in 2% of patients.

Postoperative course

The majority of patients (84%) underwent only 1 additional procedure following wound dehiscence. Thirteen patients required 2 or more additional operations, including 1 patient who underwent a total of 6 subsequent procedures during a single hospitalization. The median length of stay was 26 days, and the overall mortality rate was 11% (Table 1).

False positive analysis

Of 15 false positives, 7 (47%) were patients in which the fascia was intentionally left open during the initial procedure. Three patients (20%) returned to the operating room because of concern for postoperative wound dehiscence, but only a superficial fluid collection without fascial dehiscence was involved. An additional 3 false positives (20%) were patients in which a second procedure was performed for a reason other than wound dehiscence (eg, postoperative hemorrhage, planned reinternalization of a ventriculoperitoneal shunt). Finally, 2 operations addressed the sequelae of prior abdominal surgical interventions (eg, colostomy takedown, ventral hernia repair) but were incorrectly coded (Table 2).

Table 2. Characteristics of False Positives

	False positives (n = 15), n (%)
Abdomen intentionally left open	7 (47)
Superficial fluid collection	3 (20)
Other reason (eg, bleeding)	3 (20)
Addressed sequelae of prior surgery	2 (13)

DISCUSSION

We found a PPV of 87% for PSI 14 within the VHA system. This is higher than the PPV of 72% previously reported for PSI 14 when comparing VHA administrative data with chart-abstracted data derived from the VA's National Surgical Quality Improvement Program.[10] It is also higher than the reported PPV range of 34% to 75% for the pediatric postoperative wound dehiscence indicator (pediatric quality indicator 11).[11] The PPV of PSI 14 also compares favorably with the PPVs of other PSIs. It has the highest PPV among all PSIs evaluated in the VA system.[9,12,13] It is also superior to the published PPVs of all other PSIs validated in the private sector other than PSI 15, accidental puncture and laceration, which had a PPV of 91%.[14-19]

The relative value of a PPV of 87% primarily depends on its intended use. Eighty-seven percent may be considered sufficiently high when broadly identifying opportunities for patient safety improvement. Alternatively, 87% is arguably unsatisfactory if the data are being used to determine pay-for-performance parameters or for disciplinary measures. It is within this latter context that an analysis of false positives is instructive, both as a reminder of the limitations of ICD-9-CM–based triggers and as a means of improving the PSI.

Our review of false positive cases highlighted potential sources of coding error. Seven, or 47%, of the false positives were patients in which the fascia was intentionally left open owing to intra-abdominal contamination or concern for development of compartment syndrome. Clinical descriptors that highlighted the intent not to close the abdomen were apparent, in the indications for surgery, postoperative diagnosis, or operative note. For instance, one operative report noted that the "abdominal fascia was intentionally left open following the first surgery." Similarly, a second report stated, "during the first operative surgery the fascia was not closed." However, these 7 false positives were each inappropriately coded as ICD-9-CM 54.61, "reclosure of postoperative disruption of abdominal wall." Although the patients did have a subsequent operation to close the abdomen, this procedure was a planned takeback, not the result of a postoperative wound dehiscence.

This coding error may be due in part to the fine level of distinction required to correctly assign an ICD-9-CM

code. Although there is no specific code for "delayed primary closure," the 7 false positives incorrectly coded as ICD-9-CM 54.61 should have been coded as ICD-9-CM 54.62, "delayed closure of granulating abdominal wound." A trained medical coder should be able to differentiate the 2 situations. Although this clinical scenario is relatively uncommon in a typical VA medical center, failure to differentiate a postoperative wound dehiscence from an intentional open abdomen could prove troublesome if PSI 14 is evaluated within hospitals at which this practice is more commonplace, such as a trauma center. This suggests a need for increased attention to coding practices and better training of coders.

An additional 20% of cases were associated with a fluid-draining wound infection or seroma requiring operative drainage. These patients were also incorrectly coded as ICD-9-CM 54.61, instead of ICD-9-CM 86.04, "other incision with drainage of skin and subcutaneous tissue." In all cases, the fascia remained intact, although wound dehiscence was initially suspected in each case. Because fluid leakage may precede wound dehiscence, its presence in the face of a second operation understandably serves as a source of confusion. In our detailed chart analysis, clinicians filtered through the operative report to eventually determine that the fascia remained intact. This level of analysis may be untenable for the average medical coder. Moreover, some clinicians use "wound dehiscence" to signify that the superficial portion of the wound involving subcutaneous tissue is open. Although not a reason for false positives in this study, this nonspecificity of clinical terms may further confound abstractors' ability to correctly code for postoperative wound dehiscence. Ultimately, this coding error may be hard to avoid without specialized instruction.

Despite these coding errors, the PPV of PSI 14 is the highest of any PSI validated in the VA setting. This is largely because other VA-validated PSIs had a high percentage of false positives from diagnoses that were actually present on admission but not coded as such because of the lack of a present on admission code in the VA administrative database. Wound dehiscence is more likely to occur within the immediate postoperative period, as opposed to other PSIs, such as pressure ulcer, which may develop after a patient has been discharged from the hospital or sometime prior to admission.

Although the PPV of PSI 14 is relatively high, it is also important to consider whether it can identify clinically significant opportunities for quality improvement.

Several descriptive characteristics of our true positives suggest that this is the case. Within this group, the median length of stay was nearly 1 month (26 days) and 11% of patients died. These data are consistent with previously published reports noting that hospitalizations associated with postoperative wound dehiscence were associated with increased morbidity and mortality (eg, VA patients hospitalized with this PSI experienced approximately 4 times the number of in-hospital deaths, nearly 4-fold greater median length of stay, and 3 times increased cost compared with similar hospitalizations without this indicator).[20]

Although we are limited by the retrospective nature of this study, analysis of the true cases of postoperative wound dehiscence suggests that at least some cases may be preventable. Twenty-three percent of patients had a preoperative albumin level of less than 3 g/dL, suggesting the opportunity to improve perioperative nutrition in elective cases. Twenty-eight percent of patients had a diagnosis of COPD, another known risk factor for wound dehiscence. Ensuring that care of patients with COPD includes both preoperative inhalers and postoperative vigilance with pulmonary toilette may also help reduce the risk of this complication. Moreover, 9% of true positives had a wound infection. This highlights the importance of minimizing surgical site infections by using proper surgical techniques and implementing surgical infection prevention measures. Also striking is that physicians-in-training performed nearly 70% of fascial closures among patients with wound dehiscence. It is not uncommon for attending surgeons to scrub out after the critical part of the procedure and leave fascial closure to the residents. Increased vigilance by the attending surgeon during closure may reduce rates of postoperative dehiscence.

Study limitations

There are some inherent limitations to our study. We are unable to report on sensitivity, specificity, or negative predictive value of this PSI because we did not abstract the charts of patients not flagged by the PSI 14 algorithm. Future research efforts should examine these issues because the PSI may be missing clinically significant events. Additionally, using the medical record as the gold standard to determine whether the PSI coding algorithm identifies true events relies on the accuracy and completeness of physician documentation. We also did not collect information on surgical infection prevention measures; therefore, we do not know the extent to which these contributed to wound infections. Finally, this was a retrospective observational study with potential for selection, observation, and confounding biases.

CONCLUSIONS

PSI 14 has good predictive ability to identify true instances of postoperative wound dehiscence in the VA system. Its predictive value could be improved by eliminating coding

errors, such as those related to cases in which the patient's fascia was intentionally left open. This PSI identifies clinically significant events representing opportunities for real quality improvement and highlights perioperative conditions and patient risk factors that may be optimized to decrease the incidence of postoperative wound dehiscence. As it currently stands, PSI 14 appears to be a promising measure for use beyond quality improvement and screening, particularly if efforts to decrease coding mistakes are implemented; however, use for pay-for-performance may be premature.

Author Contributions

Study conception and design: Borzecki, Shin, Itani, Rosen
Acquisition of data: Cevasco, Borzecki, McCluskey, Chen, Shin, Rosen
Analysis and interpretation of data: Cevasco, Borzecki, McCluskey, Chen, Shin, Itani, Rosen
Drafting of manuscript: Cevasco, Borzecki, McCluskey, Chen, Shin, Itani, Rosen
Critical revision: Cevasco, Borzecki, Itani, Rosen

Acknowledgment: The authors would like to thank Julie Lynch, RN, and Kristine Ruggiero, RN, for their assistance with abstraction instrument and guidelines modification and data abstraction; Haytham Kaafarani, MD, MPH, for his assistance with abstraction instrument and guidelines development; Dijon Fasoli, RN, PhD, for conducting IRR assessment; and Hillary Mull, MPP, for her technical assistance in developing the electronic abstraction form.

REFERENCES

1. Franz MG, Steed DL, Robson MC. Optimizing healing of the acute wound by minimizing complications. Curr Probl Surg 2007;44:691–763.
2. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA 2003;290:1868–1887.
3. Carlson MA. Acute wound failure. Surg Clin North Am 1997; 77:607–638.
4. van Ramshorst GH, Nieuwenhuizen J, Hop WC, et al. Abdominal wound dehiscence in adults: development and validation of a risk model. World J Surg 2010;34:20–27.
5. Webster C, Neumayer L, Smout R, et al. Prognostic models of abdominal wound dehiscence after laparotomy. J Surg Res 2003; 109:130–137.
6. Riou JP, Cohen JR, Johnson H. Factors influencing wound dehiscence. Am J Surg 1992;163:324–330.
7. Agency for Healthcare Research and Quality. AHRQ Quality Indicators: Guide to Patient Safety Indicators. Version 3.0a. Rockville, MD; 2003. Version 3.1; 2007.
8. Centers for Medicare & Medicaid Services. Reporting hospital quality data for annual payment update. Available at: https://www.cms.gov/HospitalQualityInits/08_HospitalRHQDAPU.asp. Accessed September 28, 2010.
9. Kaafarani HMA, Borzecki AM, Itani KMF, et al. Validity of selected patient safety indicators: opportunities and concerns. J Am Coll Surg 2011;212:924–934.
10. Romano PS, Mull HJ, Rivard PE, et al. Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data. Health Serv Res 2009;44: 182–204.
11. Scanlon MC, Harris JM, Levy F, Sedman A. Evaluation of the agency for healthcare research and quality pediatric quality indicators. Pediatrics 2008;121:e1723–e1731.
12. Borzecki AM, Kaafarani HMA, Cevasco M, et al. How valid is the patient safety indicator "postoperative hemorrhage or hematoma"? J Am Coll Surg 2011;212:946–953.
13. Borzecki AM, Kaafarani HMA, Utter GH, et al. How valid is the AHRQ patient safety indicator "postoperative respiratory failure"? J Am Coll Surg 2011;212:935–945.
14. Gallagher B, Cen L, Hannan E. Validation of AHRQ's patient safety indicator for accidental puncture or laceration. In: Henriksen K, Battles J, Marks E, et al, eds. Advances in Patient Safety: From Research to Implementation. Concepts and Methodology. Vol 2. Rockville, MD: Agency for Healthcare Research and Quality; 2005. AHRQ Publication No. 05-0021-2.
15. White RH, Sadeghi B, Tancredi DJ, et al. How valid is the ICD-9-CM based AHRQ patient safety indicator for postoperative venous thromboembolism? Med Care 2009;47:1237–1243.
16. Utter GH, Zrelak PA, Baron R, et al. Positive predictive value of the AHRQ accidental puncture or laceration patient safety indicator. Ann Surg 2009;250:1041–1045.
17. White RH, Garcia M, Sadeghi B, et al. Evaluation of the predictive value of ICD-9-CM coded administrative data for venous thromboembolism in the United States. Thromb Res 2010;126:61–67.
18. Utter GH, Cuny J, Sama P, et al. Detection of postoperative respiratory failure: how predictive is the Agency for Healthcare Research and Quality's patient safety indicator? J Am Coll Surg 2010;211:347–354, e1-29.
19. Zrelak PA, Sadeghi B, Utter GH, et al. Positive predictive value of the AHRQ patient safety indicator for central line related-bloodstream infection ("selected infections due to medical care"). J Healthc Q 2011;33:29–36.
20. Rosen AK, Rivard P, Zhao S, et al. Evaluating the patient safety indicators: how well do they perform on Veterans Health Administration data? Med Care 2005;43:873–884.

Validity of the AHRQ Patient Safety Indicator "Central Venous Catheter-Related Bloodstream Infections"

Marisa Cevasco, MD, MPH, Ann M Borzecki, MD, MPH, William J O'Brien, MS, Qi Chen, MPH, Marlena H Shin, JD, MPH, Kamal MF Itani, MD, FACS, Amy K Rosen, PhD

BACKGROUND: "Central venous catheter-related bloodstream infections" (CR-BSIs) is one of the patient safety indicators (PSI 7) developed by the Agency for Healthcare Research and Quality (AHRQ) to screen for potential safety events. We sought to investigate the validity of this PSI using the medical record as the gold standard.

STUDY DESIGN: We conducted a retrospective cross-sectional study of all hospitalization records that met the criteria for PSI 7 within Veterans Health Administration (VA) hospitals from fiscal years 2003 to 2007. Trained abstractors used a standardized abstraction tool to review electronic medical records for the presence of a CR-BSI and the clinical circumstances surrounding the event. We determined the validity of this PSI by calculating its positive predictive value (PPV), and analyzed both true and false positive cases.

RESULTS: Of 112 reviewed cases, 42 were true events of CR-BSIs, yielding a PPV of 38% (95% CI 29% to 47%). Seventy cases were false positives; these were attributed to correct ICD-9-CM codes but had diagnoses that fell outside the scope of the indicator (n = 28, 40%), coding inaccuracies (n = 21, 30%); and present on admission (POA) diagnoses (n = 21; 30%). Among the 42 patients with CR-BSIs, catheters were left in place for an average of 11 days, and 20% (n = 8) were placed in the femoral position.

CONCLUSIONS: PSI 7 has relatively poor predictive ability for identifying true events. Coding-related issues were the main reason for the low PPV. Implementing POA codes and using more specific ICD-9-CM codes would improve its validity. As it currently stands, PSI 7 should not be used as a pay-for-performance measure, but should be limited to use in internal quality improvement efforts. (J Am Coll Surg 2011;212:984–990. © 2011 by the American College of Surgeons)

Central venous catheter-related bloodstream infections (CR-BSIs) are associated with significant morbidity and mortality, prolonged hospitalization, and increased medical costs. They have an attributable mortality rate as high as 25%, are associated with up to a 14-day increase in length of stay, and an incremental in-hospital cost of $45,000 per patient.[1-5] CR-BSIs affect more than 80,000 patients per year in the United States alone.[6]

Vascular catheter-related infections are preventable.[6] Evidence-based processes of care for preventing CR-BSIs include proper hand hygiene, use of maximal barrier precautions, and chlorhexidine gluconate for insertion site preparation.[5,7] Daily review of catheter necessity, prompt removal of unnecessary lines, and avoiding the femoral vein as the insertion site are also associated with a decreased rate of infection.[5,7] One notable initiative at the University of Michigan Health and Hospital Association (MHA) Keystone ICU Project nearly eliminated CR-BSIs by implementing these measures. They decreased baseline rates of CR-BSIs from a mean of 7.7 and median of 2.7 per 1,000 catheter days, to 1.3 and 0, respectively, at 18 months postimplementation.[6] Significantly, 3 years later these rates remained at a mean of 1.1 and a median of 0.[6]

Disclosure Information: Nothing to disclose.

Received December 9, 2010; Revised January 31, 2011; Accepted February 1, 2011.

From the Department of Surgery (Cevasco, Itani) and the Center for Organization, Leadership, and Management Research (O'Brien, Chen, Shin, Rosen), VA Boston Healthcare System; the Department of Surgery, Brigham and Women's Hospital (Cevasco); the Department of Health Policy and Management, Boston University School of Public Health (Borzecki, Chen, Rosen); Boston University School of Medicine (Borzecki, Itani); and Harvard Medical School (Itani), Boston, MA; and the Center for Health Quality, Outcomes and Economic Research, Bedford VAMC, Bedford, MA (Borzecki).

Correspondence address: Amy K Rosen, PhD; Center for Organization, Leadership, and Management Research (COLMR), VA Boston Healthcare System, 150 S Huntington Ave (152 M), Boston, MA 02130. email: akrosen@bu.edu

ISSN 1072-7515/11/$36.00
doi:10.1016/j.jamcollsurg.2011.02.005

Abbreviations and Acronyms

AHRQ = Agency for Healthcare Research and Quality
CR-BSI = central venous catheter-related bloodstream infection
IRR = inter-rater reliability
PICC = peripherally inserted central venous catheter
POA = present on admission
PPV = positive predictive value
PSI = patient safety indicator
VA = Veterans Health Administration

Because CR-BSI is a significant complication of inpatient hospitalization that may be prevented through specific processes of care, the Agency for Healthcare Research and Quality (AHRQ) chose it as a patient safety indicator (PSI). The PSIs, first released in March 2003, were designed to identify areas of concern and potential preventability that occur during inpatient admissions. The free PSI software provides hospitals with a way to assess the incidence of adverse events that occur during inpatient hospitalizations using administrative data (ICD-9-CM codes) found in the discharge record. Currently, there are 20 hospital-level PSIs that cover a spectrum of surgical, medical, and obstetric adverse events.[8] PSI 7, previously known as "selected infections due to medical care," was renamed "central venous catheter-related bloodstream infections" to clarify the intent of the indicator.[9]

The PSIs were developed for internal review and quality improvement projects, but are increasingly used to assess hospital performance, compare facilities, and as pay-for-performance measures.[8,10] More than 100 organizations use PSIs for public reporting and hospital profiling. A very similar measure, vascular catheter-associated infection, is included among the list of hospital-acquired conditions for which hospitals will no longer receive reimbursement from Centers for Medicare and Medicaid Services (CMS).[11] Despite these recent policies, there are currently no data on the validity of this PSI or its related hospital-acquired conditions measure. Moreover, this policy was enacted despite studies of other PSIs showing wide variability in positive predictive values (PPVs), a measure of their accuracy against the gold standard of medical record review.[12-17]

As part of a larger study, we investigated the PPV of PSI 7 within the Veteran's Health Administration (VA). This is the first study, to our knowledge, that investigates the PPV of this PSI using the medical record as the gold standard. We also investigated the types of events incorrectly captured by the PSI algorithm in order to better understand the reasons behind its failure. Finally, we sought to gain insight into the circumstances surrounding true events of CR-BSI, with the goal to recommend improvements in processes of care.

METHODS

Study design

This was a retrospective chart review based on a computerized random sample of VA administrative data that met criteria for PSI 7. AHRQ software version 3.1a was applied to fiscal year 2003 to 2007 (October 1, 2002 to September 30, 2007) inpatient administrative data to identify patients coded as having a CR-BSI. Institutional Review Board approvals were obtained from the VA Boston Healthcare System and the Bedford VA Medical Center.

Definition of PSI 7

PSI 7, central venous catheter-related bloodstream infections, included all medical or surgical discharges of patients aged 18 years or older with an ICD-9-CM code of 999.3(infection following infusion, injection, transfusion or vaccination); or 996.62 (infection or inflammatory reaction to a vascular device, implant or graft, in any secondary diagnosis field). Discharges were excluded if the length of stay was less than 2 days, if the patient was immunocompromised, or had a cancer diagnosis. Discharges were also excluded if ICD-9-CM codes 999.3 or 996.62 were found in the principal diagnosis field, suggesting that the condition was present on admission (POA).[8]

Hospital selection

Twenty-eight VA hospitals were randomly selected from the network of 158 acute care VA hospitals. This ensured sufficient exposure to interfacility coding variability while maintaining a manageable number of hospitals for individual medical record review. Our hospital sample included facilities with a range of observed and expected patient safety events. Details of our sampling scheme have been previously published.[13]

Case selection

Four medical records flagged with PSI 7 were randomly selected from each of the 28 hospitals, to generate a sample size of 112 medical records. One hundred twelve cases were selected to ensure reasonably narrow PPV confidence intervals (± 10% to 20%).

Medical record abstraction

Two trained nurses conducted a retrospective chart review of the 112 cases. Standardized chart abstraction tools and guidelines developed by AHRQ for "selected infections due to medical care" were modified to be specific to CR-BSI; these were then adapted to the VA's electronic medical

record and designed to generate inter-rater reliability (IRR) measurement standards of greater than or equal to 90% observed agreement. The tool also directed the nurses' chart review, and facilitated identification of both true and false positive cases.

Nurse training included several sessions discussing the rationale of PSI 7, the likely sources of information needed from the electronic medical record, and a systematic chronology for chart abstraction. We also conducted IRR testing to ensure standardized and reliable abstraction; this was performed at the start, and later in abstraction to rule out abstractor drift. The nurses achieved 91% observed agreement after an initial round of 5 records. Observed agreement on late IRR testing on 5 records was similarly 91%. Further details of the chart abstraction are described elsewhere.

Analysis

Positive predictive value (PPV)

Medical records were reviewed to detect the occurrence of a CR-BSI. PPV was calculated as the rate of true positives divided by the total number of medical records reviewed. We also derived 95% confidence intervals around that estimate.

True positive analysis

We performed descriptive analyses of all patients with confirmed CR-BSIs as documented in the medical record. Specifically, we relied on clinician documentation of an infection from a central venous catheter, such as a triple lumen, Hickman, peripherally inserted central catheter (PICC), or portacath. We excluded infections from peripheral venous catheters, arterial lines, or vascular grafts. We also investigated multiple continuous and categorical variables including demographic characteristics, comorbidities, site of central line placement, indication for placement, diagnosis of CR-BSI, and patient outcomes. All statistical analyses were performed using SAS version 9.1 (SAS Institute, Inc).

False positive analysis

We performed detailed analysis of all cases that were not true instances of CR-BSI to understand the reason behind their inappropriate capture by the PSI 7 algorithm.

RESULTS

From the 28 selected hospitals, 467,434 cases met the eligibility criteria for PSI 7, and 1,033 cases were flagged as having a CR-BSI. This yields an observed rate of 2.2 per 1,000 discharges, comparable to a rate of 2.0 per 1,000 discharges in the overall VA population. Of these 1,033 cases, 112 were chosen for detailed medical record review as described above. Forty-two cases were true cases of CR-BSI, yielding a PPV of 38% (95% CI 29% to 47%). The remaining 70 cases were false positives.

Table 1. Characteristics of True Positives

Variable	True positives n=42
Demographics and comorbid conditions	
Male gender, n (%)	41 (99)
Age, y, mean	61 ± 10
Diabetes, n (%)	17 (41)
Systemic steroids, n (%)	6 (14)
Presenting conditions, n (%)	
Urinary tract infection	8 (19)
Pneumonia	5 (12)
Gastroenteritis	4 (10)
Pancreatitis	3 (7)
Cellulitis	3 (7)
Clostridium difficile infection	2 (5)
Other*	3 (7)

*Includes 1 patient with a small bowel perforation, 1 with bacterial meningitis, and 1 with a surgical site infection.

True positive analysis

Patients were 98% male (n = 41) with a mean age of 60.6 ± 10.3 years. Seventeen patients (41%) had diabetes, and 6 (14%) patients had been on intravenous or oral steroids within the previous month (Table 1). None of the other patients had received any other type of immunosuppressive drug within the past month and none of the patients underwent radiation therapy.

A total of 23 patients (55%) were noted in the medical record as having 1 or more infectious or inflammatory conditions POA. Specifically, 15 patients had 1 such condition and 8 were noted as having 2 or more. Eight patients had a urinary tract infection POA, 5 were admitted with pneumonia, and 4 patients had gastroenteritis. Three patients were admitted with pancreatitis, 3 had cellulitis, and 2 were admitted with a *Clostridium difficile* infection. One patient was admitted with a small bowel perforation, 1 with bacterial meningitis, and 1 with a surgical site infection (Table 1). (Of note, an infection POA does not exclude a case from being reviewed for a CR-BSI per the definition of PSI 7).

Of the 3 patients who had central lines POA, these were all peripherally inserted central venous catheters (PICCs) for administration of total parenteral nutrition. None of these 3 patients had blood cultures drawn on admission because there was no suspicion of infection. Two of these PICCs were subsequently considered to be the source of the CR-BSI that developed during the hospitalization; the third patient had a second line placed that was considered to be the source.

Table 2. Characteristics of Catheter-Related Bloodstream Infections

Characteristic	True positives (n=42)
Infectious organism, n (%)	
Gram-positives	
Methicillin-sensitive *Staphylococcus aureus*	12 (29)
Methicillin-resistant *Staphylococcus aureus*	6 (14)
Coagulase-negative *Staphylococcus* species	5 (12)
Enterococcus species	5 (12)
Vancomycin-resistant *Enterococcus faecalis*	1 (2)
Gram negatives and *Candida*	
Eschericia coli	3 (7)
Pseudomonas aerigunosa	2 (5)
Candida species	3 (7)
Source of blood culture, n (%)	
Peripheral	17 (41)
Central	14 (33)
Not specified	11 (26)
Site of catheter placement, n (%)	
Internal jugular vein	12 (29)
Subclavian vein	10 (24)
Femoral vein	8 (19)
Peripherally inserted central venous catheter	12 (29)
Emergent or urgent lines, n (%)	14 (43)
Catheter tip, n (%)	
Sent for culture	30 (71)
Positive for organisms	16 (38)
Blood cultures sent within 48 h of central line placement, n (%)	4 (11)
Lines changed over a wire, n (%)	3 (16)
Duration catheters in place, mean, d	11 ± 9
Symptoms or clinical signs, n (%)	
Fever greater than 38.0°C	36 (86)
Leukocytosis or bandemia*	20 (48)
Rigors or chills	10 (24)
Hypotensive†	8 (19)
Transfer to higher level of care, n (%)	7 (17)
Death attributed in part to CR-BSI, n (%)	6 (14)

*Leukocytosis defined as the presence of a white blood cell count greater than 12,000; bandemia defined as greater than 10% band forms.
†Documented systolic blood pressures of less than 90 mmHg.
CR-BSI, catheter-related bloodstream infection.

All 42 true positive cases had positive blood cultures drawn after placement of their central line. The majority of infections were due to gram-positive organisms (Table 2). Seventeen patients had positive blood cultures (41%) obtained from peripheral lines, and 14 patients had positive blood cultures (33%) drawn from the central line; of this latter group, 7 patients had blood cultures drawn from both central and peripheral sources. In the remaining 11 patients (26%) with positive blood cultures, the source of the sample was not documented. In addition, 30 patients had catheter tips sent for culture, of which 16 (53%) were positive for infection. Eleven percent of all true cases (n = 4) had positive blood cultures that were sent within the first 48 hours of central line placement (Table 2).

Of the catheters associated with infection, 12 (29%) were placed in the internal jugular vein, 12 (29%) were PICCs (including the 2 PICCs that were POA), 10 (24%) were in the subclavian position, and 8 (19%) were in the femoral vein. Of the 41 patients for whom documentation was available, 19 (46%) had the infected catheter removed and a new catheter placed the same day. Eighteen catheters (43%) were placed in an urgent or emergent situation; 6 (33%) of these were femoral venous catheters. Catheters were in place for a mean duration of 11 days (median 11 days; range 2 to 56 days) (Table 2).

All patients manifested at least 1 symptom or laboratory indicator of a CR-BSI, such as a fever greater than 38.0°C, leukocytosis or bandemia, rigors or chills, or hypotension with systolic blood pressures of less than 90 mmHg. CR-BSIs resulted in 7 patients being transferred to a higher level of care, such as an ICU. Six patients' deaths were attributed in part to a CR-BSI (Table 2).

False positive analysis

Seventy patients were false positive. Of these, 40% (n = 28) of patients received a correct ICD-9-CM code, but this code did not capture the clinical intent of the indicator. Specifically, 19 patients (27%) were diagnosed with thrombophlebitis, 5 patients (7%) had an infected vascular polytetrafluoroethylene femoral-popliteal bypass graft, and 4 (6%) had cellulitis at a peripheral IV site, all of which would be coded for using ICD-9-CM code 996.62. An additional 30% (n = 21) were inaccurately coded as having a CR-BSI. Reasons for inaccurate coding included negative work-up for a CR-BSI (n = 12, 17%), or no documentation of central line, no evidence of infection or inflammation at an injection site, or infection or inflammation of a vascular device (n = 9, 13%). Finally, 30% (n = 21) of patients had a CR-BSI POA (Table 3).

DISCUSSION

The PPV of PSI 7 is 38%, indicating a poor ability to identify CR-BSIs. Recently evaluated PSIs have demonstrated PPVs between 44% and 91%, so PSI 7 ranks lowest among all PSIs evaluated to date.[12-17]

Importantly, 40% of false positives were cases that were coded correctly but represented diagnoses that were not clinically sensitive to the intent of the indicator. Specifically, there were 19 cases of superficial thrombophlebitis

Table 3. Reasons for False Positives

Reason	False positives (n=70)	
	n	%
Inaccurately coded, n (%)		
Negative rule-out	12	17
No documentation of central line or infection/ inflammation at IV site	9*	13
Outside of the scope of the indicator, n (%)		
Superficial thrombophlebitis	19	27
Infected vascular graft	5	7
Cellulitis at IV site	4	6
Present on admission, n (%)	21	30

*This includes 4 cases associated with a nonvascular catheter infection (urinary tract infections) and 1 case of nonvascular "injection/aspiration," ie, paracentesis.

from a peripheral IV, 5 cases of infected vascular bypass grafts, and 4 cases of cellulitis from a peripheral IV that were correctly captured under ICD-9-CM code 996.62, "infection or inflammatory reaction to a vascular device, implant or graft." These diagnoses do not represent a CR-BSI and are outside the scope of this PSI. A more specific ICD-9-CM code for CR-BSI, 993.1, has subsequently been developed for use on data captured after 2007. This code was designed to capture only an infection from central venous catheters, including triple lumens, Hickmans, PICCs, or portacaths, and excludes infections from peripheral venous catheters, arterial lines, or vascular grafts.[18] (Because our sample contained data from 2003 to 2007, we do not know how well this new code performs.)

Code 996.62 was also recently modified to exclude cases captured by 999.31.[18] In our study, if we were to eliminate those non-CR-BSI cases captured by the older versions of these ICD-9-CM codes, our PPV would become 63% (95% CI 54% to 72%), arguably still a low PPV, but relatively improved. This scenario highlights the vagueness of many administrative codes, as well as the importance of validation studies such as ours, which show the potential improvement that can be achieved by making ICD-9-CM codes more specific. This also underscores the fact that quality measures based on administrative codes should be interpreted cautiously, especially when chart validation is lacking.

Similar to many of the PSIs, diagnoses that were POA played a large role in limiting the potential usefulness of this indicator. Thirty percent (n = 21) of patients had a CR-BSI on admission, suggesting that a POA flag in the administrative data would increase the indicator's PPV. This has already been implemented in private hospitals and is in the early stages of adoption by the VA system. Introduction of a POA flag, in addition to using the updated ICD-9- CM codes, would generate a PPV of 81% (95% CI 73% to 88%), more in line with previously published results for other PSIs.

Coding inaccuracies are a final important component limiting the potential value of this indicator. Thirty percent of false positives (n = 21) were cases that were incorrectly coded. Of these, medical record review revealed that 12 had negative work-ups for CR-BSIs. Of the remaining 9 cases, there was no documentation of central line placement, no evidence of infection or inflammation at an injection site, or infection or inflammation secondary to a vascular device. It is unclear how these medical records were given ICD-9-CM codes of 999.3 or 996.62, particularly because chart review did not reveal any diagnosis (such as thrombophlebitis or infected vascular bypass graft) that would qualify them for such a code. The attention to detail required for ascertaining this may be untenable for the average medical coder, who is often charged with reviewing up to 15 inpatient medical records per day (B Cirrone, personal communication, November 22, 2010). Coders may not have the time to carefully peruse the complete medical record, progress notes may be contradictory, and discharge summaries may be incomplete or nebulous.

It is also important to note variations in the definitions of CR-BSI. The Centers for Disease Control and Prevention, in their *Guidelines for the Prevention of Intravascular Catheter-related Infections*, provides 3 different definitions of CR-BSI.[5] Two surveillance definitions were developed for use in quality of care and hospital performance measures; these include "laboratory-confirmed BSI" and "catheter-associated BSI." A third definition, "catheter-related BSI," was developed for clinical purposes only and provides the most stringent criteria, including positive cultures from the catheter tip and central and peripheral lines (Table 4).[5] Not only do the definitions of CR-BSI significantly differ, they may be used differently across hospitals and clinicians.[19] To obtain accurate rates of CR-BSIs, particularly when comparing facilities, it is important to know what definition of CR-BSI was used to arrive at the diagnosis. The ICD-9-CM codes (996.62 and 999.31) are not specific enough to differentiate between the CDC definitions of laboratory-confirmed BSI, catheter-associated BSI, or CR-BSI. Additionally, the indicator itself does not specify a definition of CR-BSI, and it may not be realistic to expect a code-based algorithm to differentiate between these 3 definitions.

A separate concern regarding this indicator is that it potentially flags events also captured under PSI 13 (postoperative sepsis) potentially penalizing hospitals twice for the same event. For example, another study described by

Table 4. Centers for Disease Control Definitions of Catheter-Related Bloodstream Infections

Definitions intended for surveillance measures
Catheter-associated BSI • Vascular access device that terminates at or close to the heart or one of the great vessels, and • BSI is considered to be associated with a central line if the line was in use during the 48-h period before development of the BSI. If the time interval between onset of infection and device use is > 48 h, there should be compelling evidence that the infection is related to the central line.

Laboratory-confirmed BSI (patients need to meet 1 of 2 criteria)
• Criterion 1: Patient has a recognized pathogen cultured from 1 or more blood cultures, and the pathogen cultured from the blood is not related to an infection at another site.
• Criterion 2: Patient has at least 1 of the following signs of symptoms: fever (>38.0° C), chills, or hypotension, and at least 1 of the following:
  • Common skin contaminant* cultured from 2 or more blood cultures drawn on separate occasions
  • Common skin contaminant cultured from at least 1 blood culture from a patient with an intravenous line, and the physician institutes appropriate antimicrobial therapy
  • Positive antigen test on blood and signs and symptoms with positive laboratory results are not related to an infection at another site.

Definition intended for clinical research measures

Catheter-related BSI
• Bacteremia/fungemia in a patient with an intravascular catheter with at least 1 positive blood culture obtained from a peripheral vein, clinical manifestations of infections (ie, fevers, chills, and/or hypotension), and no apparent source for the BSI except the catheter.
• One of the following should be present:
  • Positive culture whereby the same organism is isolated from the catheter segment and peripheral blood;
  • simultaneous quantitative blood cultures with a ≥ 5:1 ratio CVC versus peripheral;
  • differential period of CVC culture versus peripheral blood culture positivity of >2 hours.

*Common skin contaminant includes diphtheroids, *Bacillus* spp, *Propionibacterium* spp, coagulase-negative staphylococci, or micrococci.
BSI, bloodstream infection; CVC, central venous catheter.

our group found that 13 cases of postoperative sepsis were attributed to a CR-BSI.[20] In this study, 3 of those same 13 cases were captured herein, resulting in these patients' adverse events contributing to 2 distinct PSIs. (The remaining 10 sepsis cases were not captured by this study because we chose distinct samples of 112 flagged cases per PSI for review (ie, 3 cases ended up in both samples).

Finally, it is important to ascertain whether some of the true cases of CR-BSIs are preventable. Seventy-five percent of femoral lines were placed during a code or medical emergency, when access to the subclavian or internal jugular veins is likely limited by other caregivers intubating the patient or performing chest compressions. In this setting, it may be difficult to avoid cannulating the femoral vein for central access.

On the other hand, we found that the mean duration of catheter placement ultimately associated with an infectious complication was 11 days, 2 days longer than the mean duration seen in other studies. Whether this indicates that central venous catheters in our study remained sterile longer or clinicians were less vigilant about working up a CR-BSI is a source of conjecture. However, it does highlight the importance of prompt removal of nonessential catheters, particularly because the mean duration of noninfected catheterization is approximately 6 days (median 4 days; range 1 to 29 days).[21] This suggests that there may be at least some room for improvement in terms of reducing the rate of CR-BSIs and in replacing emergently placed femoral catheters as soon as possible.

Our study has several strengths. This is the first study to evaluate the accuracy of PSI 7 in identifying CR-BSI in any setting. The VA system is unique in that it is not subject to financial incentives, nor is it currently tracking PSI rates, so there are no penalties or benefits for not coding or clinically documenting PSIs. Additionally, our sample was randomly drawn from a nationally representative group of VA hospitals, ensuring broad geographic representation and diversity of inpatient clinical encounters. IRR testing occurred twice during the data abstraction process and achieved a high level of agreement on both occasions. Last, physicians (AB) were available to clarify clinical issues with the nurse-abstractors as necessary throughout the abstraction process to maximize accuracy and reliability of our findings.

There are certain limitations to our study. Our population was predominately older male veterans, which may limit its generalizability. As a retrospective observational study, there is potential for selection, observation, and confounding biases. We were not able to investigate all aspects of central line placement such as use of chlorhexidine gluconate, maximal barrier protection, or hand hygiene; this information was not commonly included in the medical record. Also, we were unable to report sensitivity, specificity, or negative predictive value because we did not investigate cases that were not flagged for a CR-BSI. Finally, we relied on the completeness and accuracy of the medical record.

Future efforts should focus on investigating the validity of this indicator in more up-to-date datasets that include recently revised ICD-9-CM codes, such as 993.1, and POA codes. It would also be instructive to investigate the rate of true cases not receiving an ICD-9-CM code for CR-BSI (ie, the negative predictive value) to get a sense of how many cases of CR-BSI are not captured by administrative codes.

CONCLUSIONS

We determined that at least in the VA system, PSI 7 has poor predictive ability, with a PPV of 38%. Implementing POA codes and using newly revised ICD-9-CM codes would improve its PPV to 81%. Nonetheless, even with a moderate predictive ability of 81%, the use of PSI 7 in pay-for-performance metrics is premature; its utility for internal quality improvement or measurement should be viewed positively. Despite this, there is some evidence that the rate of CR-BSIs in this population may be at least modestly reduced, particularly through clinician vigilance regarding duration of catheterization.

Author Contributions

Study conception and design: Itani, Borzecki, Rosen, Shin
Acquisition of data: Cevasco, Borzecki, O'Brien, Rosen, Shin, Chen
Analysis and interpretation of data: Itani, O'Brien, Cevasco, Borzecki, Rosen, Shin, Chen
Drafting of manuscript: Itani, Cevasco, Borzecki, Shin, Chen, Rosen, O'Brien
Critical revision: Itani, Cevasco, Borzecki, Rosen

Acknowledgment: The authors would like to thank Kathleen Hickson, RN, MA, and Sally MacDonald, RN for their assistance with abstraction instrument and guidelines modification, and data abstraction; and Susan Loveland, MAT for her programming expertise and assistance with identifying the hospital sample and flagged cases.

REFERENCES

1. Maki DG. Nosocomial bacteremia. An epidemiologic overview. Am J Med 1981;70:719–732.
2. Collignon PJ, Munro R, Sorrell TC. Systemic sepsis and intravenous devices. A prospective survey. Med J Aust 1984;141:345–348.
3. Velez LA, Maki DG. Analysis of risk factors for mortality in nosocomial bloodstream infection. A case–control study [Abstract]. In: Program and Abstracts of the 32nd Interscience Conference on Antimicrobial Agents and Chemotherapy, October 1992, Anaheim, CA. Am Soc Microbiol;1992:1164.
4. Pittet D, Tarara D, Wenzel RP. Nosocomial bloodstream infection in critically ill patients. Excess length of stay, extra costs, and attributable mortality. JAMA 1994;271:1598–1601.
5. O'Grady NP, Alexander M, Dellinger EP, et al. Guidelines for the prevention of intravascular catheter-related infections. MMWR Recomm Rep 2002;51(RR-10):1–29.
6. Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med 2006;355:2725.
7. Mermel LA. Prevention of intravascular catheter-related infections. Ann Intern Med 2000;132:391–402.
8. Agency for Healthcare Research and Quality (AHRQ). AHRQ Quality Indicators: Guide to Patient Safety Indicators. Version 3.0a. Rockville, MD: March 12, 2007.
9. Agency for Healthcare Research and Quality (AHRQ). AHRQ Patient Safety Indicators: Initial Software Release. Version 2.1. Rockville, MD: March 13, 2003.
10. Healthgrades patient safety methodology. Golden, CO. Available at: http://www.healthgrades.com/media/DMS/pdf/PatientSafetyMethodology.pdf. Accessed November 26, 2010.
11. United States Government Deficit Reduction Act of 2005. Section 5001(c). Washington, D.C. Available at: http://edocket.access.gpo.gov/2007/07-5801.htm. Accessed November 26, 2010.
12. Utter GH, Zrelak PA, Baron R, et al. Positive predictive value of the AHRQ accidental puncture or laceration patient safety indicator. Ann Surg 2009;250:1041–1045.
13. Kaafarani HMA, Borzecki AM, Itani KMF, et al. Validity of selected patient safety indicators: opportunities and concerns. J Am Coll Surg 2011;212:924–934.
14. White RH, Sadeghi B, Tancredi DJ, et al. How valid is the ICD-9-CM based AHRQ patient safety indicator for postoperative venous thromboembolism? Med Care 2009;47:1237–1243.
15. Gallagher B, Cen L, Hannan E. Validation of AHRQ's patient safety indicator for accidental puncture or laceration. In: Henriksen K, Battles J, Marks E, et al, eds. Advances in Patient Safety: From Research to Implementation. Concepts and Methodology. Vol 2. Rockville, MD: Agency for Healthcare Research and Quality; 2005. AHRQ Publication No. 05-0021-2.
16. Utter GH, Cuny J, Sama P, et al. Detection of postoperative respiratory failure: How predictive is the AHRQ Patient Safety Indicator? J Am Coll Surg 2010;211:347–354.
17. Zrelak PA, Sadeghi B, Utter GH, et al. Positive predictive value of the agency for healthcare research and quality patient safety indicator for central line-related bloodstream infection ("selected infections due to medical care"). J Healthcare Qual 2011;33:29–36.
18. Agency for Healthcare Quality and Research (AHRQ) National Quality Measures Clearinghouse. Available at: http://www.guideline.gov/content.aspx?id=15390. Accessed December 6, 2010.
19. Lin MY, Hota B, Khan YM, et al. Quality of traditional surveillance for public reporting of nosocomial bloodstream infection rates. JAMA 2010;304:2035–2041.
20. Cevasco M, Borzecki AB, Chen Q, et al. Positive predictive value of the AHRQ patient safety indicator postoperative sepsis: implications for practice and policy. J Am Coll Surg 2011. In press.
21. LeMaster CH, Schuur JD, Pandya D, et al. Infection and natural history of emergency department-placed central venous catheters. Ann Emerg Med 2010;56:492–497.

Impact of a New Gender-Specific Definition for Binge Drinking on Prevalence Estimates for Women

Pollyanna R. Chavez, PhD, David E. Nelson, MD, MPH, Timothy S. Naimi, MD, MPH, Robert D. Brewer, MD, MSPH

Background: Binge drinking accounts for more than half of the 79,000 deaths due to excessive drinking in the U.S. each year. In 2006, the Behavioral Risk Factor Surveillance System (BRFSS) lowered the threshold for defining binge drinking among women from ≥5 drinks to ≥4 drinks per occasion, in accordance with national recommendations.

Purpose: To assess changes in binge-drinking prevalence among women.

Methods: The relative and absolute change in binge drinking among U.S. adult women was assessed using pooled BRFSS data from the 2 years before (2004–2005) and after (2006–2007) the implementation of the new gender-specific definition. Analyses were conducted in 2008–2009.

Results: Binge-drinking prevalence among women increased 2.6 percentage points (from 7.3% in 2004–2005 to 9.9% in 2006–2007), a 35.6% relative increase. The percentage of women who reported consuming exactly 4 drinks in 2006 (3.6%) was similar to the increase in the prevalence of binge drinking among women that was observed from 2005 to 2006 (absolute change=2.9 percentage points).

Conclusions: The new gender-specific definition of binge drinking significantly increased the identification of women drinking at dangerous levels. The change in prevalence among women was primarily due to the change in the definition and not to actual changes in drinking behavior. The new gender-specific definition of binge drinking can increase the usefulness of this measure for public health surveillance and support the planning and implementation of effective prevention strategies (e.g., increasing alcohol excise taxes).

(Am J Prev Med 2011;40(4):468–471) Published by Elsevier Inc. on behalf of American Journal of Preventive Medicine

Introduction

Binge drinking is responsible for more than half of the estimated 79,000 deaths and two thirds of the 2.3 million years of potential life lost annually due to excessive alcohol use in the U.S.[1] It is also a risk factor for many health and social problems.[2] Since the Monitoring the Future Study[3] began using a 5-drink measure for high school students in 1975, most national surveys have defined binge drinking as consuming 5 or more drinks on an occasion (or in a row) for both women and men.[4,5] The Harvard School of Public Health College Alcohol Study used a gender-specific measure of ≥5 drinks for men, and ≥4 drinks for women, because of gender differences in the risk of alcohol-related harms at these levels.[6–8] The use of a 4-drink threshold for defining binge drinking in women is justified also because women generally have a smaller stature than men and because of physiologic differences that affect the absorption and distribution of alcohol (e.g., women absorb alcohol more rapidly than men).[9]

Recognizing these differences, in 2004 the National Institute on Alcohol Abuse and Alcoholism (NIAAA) Advisory Council endorsed the use of a 4-drink threshold for defining binge drinking in women.[10] The National Epidemiologic Survey on Alcohol and Related Condi-

From the Alcohol Program, Emerging Investigations and Analytic Methods Branch, Division of Adult and Community Health, National Center for Chronic Disease Prevention and Health Promotion, CDC (Chavez, Brewer), Atlanta, Georgia; the Cancer Prevention Fellowship Program, National Cancer Institute, NIH (Nelson), Bethesda, Maryland; and the Section of General Internal Medicine, Boston University School of Medicine (Naimi), Boston, Massachusetts

Address correspondence to: David E. Nelson, MD, MPH, Director, Cancer Prevention Fellowship Program, Center for Cancer Training, National Cancer Institute, 6120 Executive Boulevard, EPS Building Suite 150E, Bethesda MD 20892-7105. E-mail: nelsonde@mail.nih.gov.

0749-3797/$17.00
doi: 10.1016/j.amepre.2010.12.008

Table 1. Binge-drinking prevalence (% [95% CI]) among U.S. adults, BRFSS, 2004–2007

	2004 *n*=303,822	2005 *n*=356,112	2006[a] *n*=355,710	2007[a] *n*=430,912
Women	7.5 (7.3, 7.8)	7.1 (6.9, 7.3)	10.0 (9.7, 10.3)	9.9 (9.7, 10.2)
Men	22.7 (22.2, 23.2)	22.0 (21.5, 22.5)	20.7 (20.2, 21.2)	21.5 (21.0, 22.0)
Overall	14.8 (14.6, 15.1)	14.3 (14.0, 14.6)	15.2 (14.9, 15.5)	15.5 (15.3, 15.8)

[a]In 2006, the BRFSS changed the threshold to define binge drinking among women from consuming ≥5 drinks on an occasion to ≥4 drinks on an occasion in the past 30 days.
BRFSS, Behavioral Risk Factor Surveillance System

tions Survey also began using 4 drinks to define binge drinking among women that year,[11] and the Behavioral Risk Factor Surveillance System (BRFSS) did so in 2006. The present study assessed the impact of the new definition on BRFSS estimates of binge drinking among women and examined whether any changes in prevalence were attributable to the new definition or to actual changes in drinking behavior of women.

Methods

Data for the present study came from the 2004–2007 BRFSS (more details available at www.cdc.gov/brfss).[12] The number of respondents ranged from 303,822 in 2004 to 430,912 in 2007, and median state response rates ranged from 50.6% to 52.7%. Data analyses were conducted in 2008–2009 using SAS-callable SUDAAN. All data were weighted to produce population-based national estimates.

Current drinkers were defined as those who reported consumption of alcohol in the past 30 days. To assess binge drinking, in 2004 and 2005 current drinkers were asked: *Considering all types of alcoholic beverages, how many times during the past 30 days did you have 5 or more drinks on an occasion?* In 2006 and 2007, the definition of binge drinking varied by gender: *Considering all types of alcoholic beverages, how many times during the past 30 days did you have [5 (for men)/4 (for women)] or more drinks on one occasion?*

To assess whether changes in binge-drinking prevalence among women were due to the adoption of a 4-drink threshold, the prevalence of binge drinking among women was compared to that of men (for whom no definition change occurred). In addition, data from another BRFSS question on the maximum number of drinks consumed by women during a drinking occasion were used to assess (1) the prevalence among women of consuming a maximum of 4 or more drinks (≥4) or 5 or more drinks (≥5) in 2005 and 2006 (before and after the new binge definition for women), and (2) the prevalence of consuming exactly 4 drinks in 2006 (when the new gender-specific binge-drinking definition was implemented).

Results

There were no significant changes in binge-drinking prevalence from 2004 to 2005, or from 2006 to 2007 among men and women (Table 1). Therefore, data were pooled for 2004–2005 and 2006–2007 and the two time periods compared. Binge-drinking prevalence for women increased from 7.3% to 9.9% (absolute increase=2.6%; relative increase=35.6%) during this time, whereas no increase was seen in binge drinking among men (Table 2).

The largest absolute increases in prevalence among women were generally among those with the highest baseline levels of binge drinking, including those aged 25–34 years, whites, those with incomes ≥$75,000, college graduates, and those who were employed (Table 2). The largest relative increases in binge-drinking prevalence were among women aged ≥45 years, those with household incomes ≥$75,000, and college graduates. Among women of childbearing age (18–44 years), binge prevalence increased from 11.3% to 14.5%.

Using data from the maximum number of drinks question, the percentage of women consuming ≥4 drinks increased slightly from 9.0% (95% CI=8.8%, 9.3%) to 9.5% (95% CI=9.2%, 9.8%) from 2005 to 2006. Furthermore, the percentage of women who reported consuming exactly 4 drinks in 2006 was 3.6% (95% CI=3.4%, 3.8%), which was similar to the increase in binge-drinking prevalence observed from 2005 to 2006 (absolute change=2.9 percentage points, 95% CI = 2.7, 3.0).

Discussion

Lowering the BRFSS threshold for defining binge drinking among women from ≥5 drinks to ≥4 drinks per occasion, in accordance with national standards,[10] increased the absolute prevalence of this behavior among U.S. women by approximately 3 percentage points, which corresponded to a one-third relative prevalence increase. Given the evidence and rationale for lowering the threshold in the first place, the new binge-drinking definition improves the ability of the BRFSS to identify women drinking at levels that increase their risk and result in impairment-level blood alcohol concentrations.[10] In addition, these analyses demonstrated that the increased prevalence was attributable to the change in the binge-drinking threshold.

The larger relative increases in binge-drinking prevalence following the adoption of the new definition among women in older age groups, with higher income levels, or with more education probably reflects the greater sensitivity of this definition for indentifying women who were drinking just below the 5-drink threshold as well as the

Table 2. Prevalence of binge drinking among women, stratified by selected sociodemographic characteristics, 2004–2007

Characteristic	2004–2005[a] (% [95% CI])	2006–2007[b] (% [95% CI])	Absolute difference (percentage points)	Relative percentage difference
All women	7.3 (7.1, 7.5)	9.9 (9.7, 10.1)	2.6 (2.3, 2.9)	35.6 (31.5, 39.7)
Age (years)				
18–24	16.8 (15.9, 17.6)	19.5 (18.5, 20.6)	2.7 (1.3, 4.1)	16.1 (7.9, 24.3)
25–34	10.7 (10.3, 11.2)	14.5 (13.9, 15.1)	3.8 (3.1, 4.5)	35.5 (28.8, 42.2)
35–44	8.4 (8.1, 8.7)	11.7 (11.3, 12.1)	3.3 (2.8, 3.8)	39.3 (33.1, 45.5)
45–64	4.7 (4.5, 5.0)	7.7 (7.5, 8.0)	3.0 (2.7, 3.3)	63.8 (56.9, 70.8)
≥65	1.1 (1.0, 1.3)	2.0 (1.8, 2.2)	0.9 (0.7, 1.1)	81.8 (62.4, 101.3)
Race/ethnicity				
White, non-Hispanic	7.9 (7.7, 8.1)	10.9 (10.7, 11.1)	3.0 (2.7, 3.3)	38.0 (34.2, 41.8)
Black, non-Hispanic	5.2 (4.8, 5.7)	6.9 (6.4, 7.5)	1.7 (1.0, 2.4)	32.7 (19.2, 46.2)
Hispanic	6.1 (5.6, 6.7)	8.1 (7.5, 8.8)	2.0 (1.2, 2.8)	32.8 (19.7, 45.9)
Other[c]	8.3 (7.4, 9.3)	6.5 (5.8, 7.3)	1.8 (0.6, 3.0)	27.7 (9.2, 46.2)
Household income ($)				
<25,000	6.8 (6.5, 7.1)	7.7 (7.4, 8.1)	0.9 (0.4, 1.4)	13.2 (5.9, 20.6)
25,000 to <35,000	7.6 (7.1, 8.1)	9.4 (8.8, 10.0)	1.8 (1.0, 2.6)	23.7 (13.2, 34.2)
35,000 to <50,000	7.9 (7.5, 8.4)	10.7 (10.2, 11.2)	2.8 (2.1, 3.5)	35.4 (26.6, 44.3)
50,000 to <75,000	8.1 (7.7, 8.5)	11.3 (10.8, 11.8)	3.2 (2.5, 3.9)	39.5 (30.9, 48.1)
≥75,000	8.9 (8.5, 9.4)	13.4 (13.0, 13.9)	4.5 (3.9, 5.1)	50.6 (43.8, 57.3)
Education level				
Less than high school	5.0 (4.6, 5.4)	6.2 (5.6, 6.8)	1.2 (0.5, 1.9)	24.0 (10.0, 38.0)
High school	7.0 (6.7, 7.3)	8.8 (8.5, 9.1)	1.8 (1.4, 2.2)	25.7 (20.0, 31.4)
Some college	8.5 (8.2, 8.9)	11.2 (10.8, 11.6)	2.7 (2.2, 3.2)	31.8 (25.9, 37.6)
College graduate	7.4 (7.1, 7.7)	11.1 (10.8, 11.5)	3.7 (3.2, 4.2)	50.0 (43.2, 56.8)
Employment status				
Employed	9.2 (9.0, 9.5)	12.7 (12.5, 13.0)	3.5 (3.1, 3.9)	38.0 (33.7, 42.4)
Unemployed	9.7 (8.8, 10.6)	12.2 (11.2, 13.3)	2.5 (1.2, 3.8)	25.8 (12.4, 39.2)
Not in workforce	4.5 (4.3, 4.7)	6.1 (5.8, 6.4)	1.6 (1.3, 1.9)	35.6 (28.9, 42.2)

[a]Binge drinking among women was defined as consuming ≥5 drinks on an occasion in the past 30 days.
[b]Binge drinking among women was defined as consuming ≥4 drinks on an occasion in the past 30 days.
[c]Asian, American Indian/Alaska Native, Native Hawaiian/Pacific Islander, multiracial, other

distribution of binge-drinking intensity (i.e., the number of drinks per binge) in these populations. For example, a recent CDC study concluded that the average number of drinks per binge was lower for women than men (6.9 vs 8.3); declined with increasing age (from 9.8 to 6.4 drinks); and was lower for college graduates (6.5 drinks) and those with a household income ≥$35,000 (6.8 drinks).[13]

The current study has several limitations. BRFSS data are self-reported, and alcohol measures in particular are subject to recall bias and underestimation.[14] The 2004−2007 BRFSS surveys were land-line based, and binge drinking is more common among people who exclusively use cell phones, such as those aged 18–24 years.[15,16] The prevalence of binge drinking based on both the old and new definitions among women could not be assessed within the same year. However, indirect methods (e.g., comparison of alcohol consumption by women at the 4-drink level before and after this definitional change) strongly suggest that the increased prevalence was due to this definitional change.

These results, along with prior biological evidence of gender differences in the metabolism of alcohol,[9] suggest that changing the operational definition of binge drinking for women from 5 to 4 drinks should be considered in other settings, such as in alcohol screening protocols in primary care settings. Evidence-based prevention strategies for binge drinking, such as increasing alcohol excise taxes,[17] limiting alcohol outlet density,[18] and maintaining and enforcing age-21-years minimum legal drinking-age laws[19,20] should be widely adopted.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC.

No financial disclosures were reported by the authors of this paper.

References

1. CDC. Alcohol-Related Disease Impact (ARDI) website. apps.nccd.cdc.gov/ardi/.
2. USDHHS. 10th Special Report to the U.S. Congress on Alcohol and Health. Washington DC: USDHHS. pubs.niaaa.nih.gov/publications/10report/intro.pdf.
3. Bachman JG, Johnston LD, O'Malley PM. Smoking, drinking, and drug use among American high school students: correlates and trends, 1975–1979. Am J Public Health 1981;71:59–69.
4. Naimi TS, Brewer RD, Mokdad A, Denny C, Serdula MK, Marks JS. Binge drinking among US adults. JAMA 2003;289:70–5.
5. Miller JW, Gfroerer JC, Brewer RD, Naimi TS, Mokdad A, Giles WH. Prevalence of adult binge drinking: a comparison of two national surveys. Am J Prev Med 2004;27:197–204.
6. Wechsler H, Austin SB. Binge drinking: the five/four measure. J Stud Alcohol 1998;59:122–4.
7. Wechsler H, Dowdall GW, Davenport A, Rimm EB. A gender-specific measure of binge drinking among college students. Am J Public Health 1995;85:982–5.
8. Wechsler H, Nelson TF. Binge drinking and the American college student: what's five drinks? Psychol Addict Behav 2001;15:287–91.
9. Baraona EE, Abittan CCS, Dohmen KK, et al. Gender differences in pharmacokinetics of alcohol. Alcohol Clin Exp Res 2001;25:502–7.
10. National Institute on Alcohol Abuse and Alcoholism. NIAAA Council approves definition of binge drinking. NIAAA Newsletter 2004;3:3.
11. Caetano R, Ramisetty-Mikler S, Floyd LR, McGrath C. The epidemiology of drinking among women of child-bearing age. Alcohol Clin Exp Res 2006;30:1023–30.
12. CDC. Behavioral Risk Factor Surveillance System (BRFSS). Atlanta GA: USDHHS, CDC 2001–2007. www.cdc.gov/brfss.
13. CDC. Sociodemographic differences in binge drinking among adults—14 States, 2004. MMWR Morb Mortal Wkly Rep 2009;58:301–4.
14. Harrison L, Hughes A. Introduction—the validity of self-reported drug use: improving the accuracy of survey estimates. NIDA Res Monogr 1997;167:1–16.
15. Blumberg S, Luke J. Wireless substitution: early release of estimates based on data from the National Health Interview Survey, July–December 2006. National Center for Health Statistics 2007.www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200705.pdf.
16. CDC. Vital signs: binge drinking among high school students and adults—U.S., 2009. MMWR Morb Mortal Wkly Rep 2010;59:1274–9.
17. Elder RW, Lawrence B, Ferguson A, et al.; Task Force on Community Preventive Services. The effectiveness of tax policy interventions for reducing excessive alcohol consumption and related harms. Am J Prev Med 2010;38:217–29.
18. Campbell CA, Hahn RA, Elder R, et al.; Task Force on Community Preventive Services. The effectiveness of limiting alcohol outlet density as a means of reducing excessive alcohol consumption and alcohol-related harms. Am J Prev Med 2009;37:556–69.
19. Babor TF, Caetano R, Casswell S, et al. Alcohol and public policy: no ordinary commodity. 1st ed. New York: Oxford University Press, 2003.
20. The Guide to Community Preventive Services. Preventing excessive alcohol use. Atlanta GA: Task Force Community Preventive Services, 2009. www.thecommunityguide.org/alcohol/index.html.

Update in HIV Medicine for the Generalist

Amina A. Chaudhry, MD, MPH[1], Gail Berkenblit, MD, PhD[1], Allen L. Gifford, MD[2,3], Joseph Cofrancesco Jr, MD, MPH[1], James Sosman, MD[4], and Lynn E. Sullivan, MD[5]

[1]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [2]Department of Medicine, Edith Nourse Rogers Memorial VA Medical Center, Boston, MA, USA; [3]Department of Medicine, Boston University Schools of Public Health and Medicine, Boston, MA, USA; [4]Department of Medicine, University of Wisconsin School of Medicine, Madison, WI, USA; [5]Department of Medicine, Yale University School of Medicine, New Haven, CT, USA.

KEY WORDS: HIV/AIDS; prevention; screening.
J Gen Intern Med 26(5):538–42
DOI: 10.1007/s11606-010-1580-x

INTRODUCTION

This update examines and summarizes the most recent data on prevention, testing, and treatment of HIV infection for the general internist. Our aims were: (1) to describe the most recent data on HIV prevention; (2) to discuss the recommendations and emerging evidence for routine HIV screening, particularly in community-based settings; (3) to interpret the most recent data on initiation and selection of antiretroviral therapy; and (4) to facilitate the application of these findings to the clinical practice of the generalist.

We performed a PUBMED search of from March 2008 through April 2010, using the Medial Subject Heading (MeSH) term "HIV," limited to English language articles focusing on human subjects. Additionally, the authors each reviewed studies published between March 2008 and April 2010 in the major internal medicine and HIV journals. We also performed targeted searches using the search terms "HIV prevention" and "HIV testing." Articles were included after review by consensus among a group of experts, all practicing HIV clinicians and researchers, if they met the following criteria: (1) offered novel findings in HIV prevention, HIV testing, or initiation of antiretroviral therapy; and (2) had the potential for direct clinical relevance to the practicing generalist. We narrowed down our selection by group consensus with the goal of presenting the eight to ten most relevant papers published since March 2008.

PREVENTION

Celum et al. Acyclovir and Transmission of HIV-1 from Persons Infected with HIV-1 and HSV-2. NEJM. 2010. 362:427-439

In HIV-1 infected populations, the seroprevalence of HSV ranges from 60-90%[1], and studies suggest that HSV may increase HIV transmission. In coinfected cells, HSV proteins bind HIV and promote transcription[2–5]. HSV reactivation is associated with increased HIV levels in blood and the genital tract,[6–9] and rates of sexual HIV transmission are markedly higher from persons with genital ulcers.[10] Additionally, several randomized clinical trials (RCTs) demonstrate that anti-HSV therapy decreases plasma HIV levels.[6,11–14]

This study is an RCT designed to evaluate the effect of daily acyclovir therapy on HIV transmission.[15] The investigators enrolled HIV serodiscordant heterosexual couples from seven sites in southern Africa and seven sites in eastern Africa. For each couple, the HIV-infected partner was seropositive for HSV, had CD4 cell count ≥250 cells/mm^3, no AIDS-related conditions, no current antiretroviral therapy, and no persistent genital ulcers. The HIV-negative partner was eligible whether HSV-negative or positive. The intervention group received acyclovir 400 mg twice daily, and the control group received an identical-appearing placebo. The primary outcome was HIV incidence. HIV sequencing was used to classify the transmission as 'linked' or 'unlinked.'

There were 3,360 discordant couples included in the final analysis. In 68% of couples, the woman was HIV-infected. The median CD4 count was 462 cells/mm^3. Sixty-eight percent of HIV-negative partners had HSV-2. There were 132 new HIV infections, corresponding to an incidence of 2.7 per 100 person-years (95% CI: 2.3 to 3.2). Eighty-four linked transmissions were included in the analysis, 41 in the acyclovir group and 43 in the placebo group (HR 0.92; 95% CI: 0.60–1.41).

The Bottom Line: Suppressive doses of acyclovir given for up to 2 years did not reduce HIV transmission, despite significantly decreased HIV viremia and symptomatic genital ulcers.

Wawer et al. Circumcision in HIV-Infected Men and Its Effect on HIV Transmission to Female Partners in Rakai, Uganda: A Randomised Controlled Trial. Lancet 2009.374:229-37

The World Health Organization (WHO) recommends male circumcision as a male HIV prevention strategy[16] on the basis of several recent RCTs that reduced HIV transmission from females to their male partners[17–19].

This study enrolled 922 HIV-infected uncircumcised men aged 15–49 years of age who were randomized to receive

Received June 1, 2010
Revised August 6, 2010
Accepted November 4, 2010
Published online December 8, 2010

either immediate circumcision or circumcision delayed for 24 months[20]. One hundred sixty HIV-negative female partners were also enrolled. The primary outcome was the rate of acquisition of HIV among female partners, including all couples with at least one follow-up visit for the female partner. Seventeen of 92 (18%) women in the intervention group and 8 of 67 (12%) in the control group had incident HIV infection during the study period. Over 24 months, the cumulative probability of HIV infection was 21.7% (95% CI: 12.7–33.4) for women in the intervention group and 13.4% (95% CI: 6.7–25.8) for those in the control group. In a Cox proportional hazards regression analysis, the adjusted hazard ration (HR) was 1.49 (95% CI: 0.62–3.57; p=0.368). There were no significant differences in HIV incidence by participant characteristics or by women's self-reported risk behaviors.

The Bottom Line: The trial was stopped early because of ineffectiveness: male circumcision of HIV-infected men did not reduce transmission of HIV to female partners of HIV-infected men in this study over a 24-month period.

Van Damme et al. Lack of Effectiveness of Cellulose Sulfate Gel for the Prevention of Vaginal HIV Transmission. NEJM. 2008. 359:463-472

More than half of all adults living with HIV/AIDS in sub-Saharan Africa are women.[21] Most strides that have been made in HIV prevention (condoms, circumcision) depend largely on male cooperation. Topical microbicides offer the possibility of initiation by women. Cellulose sulfate is an entry inhibitor with in-vitro activity against HIV and demonstrated safety and tolerability. This was a randomized, double-blind, placebo-controlled trial[22] of cellulose sulfate gel for prevention of HIV.

Women who were 18 or older, had a negative HIV-antibody test and reported three or more acts of vaginal intercourse/week and three or more different partners in the previous 3 months were recruited from five sites in Africa and India. The intervention group received 6% cellulose sulfate gel and controls an identical placebo (the pH of the compounds were different). The primary outcome was incident HIV infection.

Among 1,398 women, at the prespecified interim analysis point, there were 24 new HIV infections in the cellulose sulfate group and 11 among those receiving placebo (HR: 2.23, 95% CI: 1.05–5.03, p=0.02), prompting early termination of the trial. While this suggested increased risk of HIV transmission among women using cellulose sulfate, the interim analysis did not include an additional six incident cases that had not been entered in the database. Analysis including these cases yielded HR: 1.61 (95 CI: 0.86–3.01, p=0.13). An additional pre-planned analysis censoring data from participants who interrupted cellulose or placebo use (most often due to pregnancy) yielded a HR 2.02 (95% CI 0.97–4.18, p=0.05).

In summary, there was a higher incidence of HIV in the cellulose sulfate group: however, this did not reach significance in the primary effectiveness analysis. It was noted by the investigators that there were non-differential pregnancy rates in the two groups. Given the contraceptive profile of cellulose sulfate, non-adherence might have been a factor in the observed results.

The Bottom Line: Cellulose sulfate is not effective in the prevention of vaginal HIV transmission.

Rerks-Ngarm et al. Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand. N Engl J Med. 2009. 361:2209-2220

This was a multicenter community-based, randomized double-blind, placebo-controlled trial in Thailand to evaluate the efficacy of vaccines to prevent HIV.[23] The vaccine protocol consisted of a primer of a recombinant canarypox vector vaccine (ALVAC-HIV) (four injections) and two booster injections of a recombinant glycoprotein 120 subunit vaccine (AIDSVAX B/E). A total of 16,402 HIV-uninfected persons aged 18–30 years were enrolled. HIV testing was performed at baseline, 24 weeks, 26 weeks, and every 6 months for 3 years of follow-up. The primary outcomes were HIV infection and early HIV viremia.

There were 132 new HIV infections, 56 in the vaccine group and 76 in the placebo group over a total of 52,985 person-years of follow-up in the intention-to-treat analysis. This result translated to an observed efficacy of 26.4% (95% CI: -4.0–47.9, p=0.08). A modified intention-to-treat analysis excluded seven persons found to be HIV-infected at baseline. In this group, there were 125 new HIV infections, 51 in the vaccine group and 74 in the placebo group over 52,985 person-years of follow-up, corresponding to an observed vaccine efficacy of 31.2% (95% CI: 1.1–52.1, p=0.04). A per-protocol additional analysis included only those who received all scheduled vaccinations in the series, maintained eligibility in the study, and had not acquired HIV by the fourth vaccination. Among this group, there were 86 new HIV infections, 36 in the vaccine group and 50 in the placebo group in 36,720 person-years of follow-up corresponding to an observed vaccine efficacy of 26.2% (95% CI: -13.3–51.9, p=0.16).

The Bottom Line: While the intention to treat analysis demonstrated only a trend towards efficacy, after excluding those with HIV at baseline, there was a significant reduction in HIV incidence in the vaccination group. Future studies must address the immune mechanisms involved as well as whether vaccine efficacy varies over time and in certain populations.

EXPANDING HIV TESTING

Bokhour et al. Barriers and Facilitators to Routine HIV Testing in VA Primary Care. J Gen Intern Med. 2009 24:1109–14

Implementing expanded HIV testing in primary care requires an understanding of the context into which broad testing will be introduced. Bokhour and colleagues conducted focus groups with patients (N=28) and health-care providers (N=13) from primary care clinics at two US Veterans Administration facilities to better understand issues in the expansion of HIV testing.[24] Patients in the four focus groups were HIV-negative men aged 35–88 years, predominantly low income, with a range of educational backgrounds. The two provider focus groups consisted of men (N=6) and women (N=7), and included physicians, nurse practitioners, and a registered nurse.

Semi-structured group interviews led to several recurrent themes. Both patients and health-care providers felt that HIV testing should be routine. Many felt that routine testing could reduce stigma. Concerns were expressed about special written consent forms and the anxieties that they may cause. Though "normalizing" the HIV testing was a consistent theme, results showed that patients wanted to be aware that HIV testing was being done and that it should be explicitly offered, with the option to decline. They also pointed out the need for clear and prompt communication of results.

The Bottom Line: Veteran patients are likely to accept routine testing, but attention to issues of stigma and clear communication about results are important for both patients and providers. To help with patient-centered discussions, the authors developed "Six R's" for routine testing: (1) Raise the topic; (2) Reassure the patient that the offer is routine; (3) provide Rationale for the test; (4) Respond to questions; (5) Request the test; and (6) tell the patient when and how they will get Results.

Arbelaez et al. Emergency Provider Attitudes and Barriers to Universal HIV Testing in the Emergency Department. J Emerg Med. 2009 Oct 13. (Epub ahead of print)

The Centers for Disease Control and Prevention (CDC) HIV testing guidelines include emergency departments (EDs) as target locations for expanded routine HIV testing.[25] This research study surveyed ED providers before and 6 months after institution of an ED-based HIV testing program to understand attitudes toward testing and how they might change with experience.[26]

One hundred eight providers (43% nurses, 29% resident physicians, 17% attending physicians, 7% nursing assistants, and 4% physician assistants) completed both pre- and post-surveys. Before starting the HIV testing intervention in the ED, many providers identified barriers to HIV testing such as inadequate resources (70%), inadequate time (51%), and concerns about assuring follow-up (50%). After experience with the intervention, two of these barriers were more frequently reported than at baseline: inadequate time (62%) and follow-up care (59%). In multivariate modeling, female providers, providers who felt they had sufficient time to test, and providers reporting sufficient legal understanding of testing issues were more likely to favor HIV testing in the ED. While most providers favored HIV testing in general (86%), fewer supported doing it in the ED (56%), and still fewer expressed willingness to offer it themselves (37%).

The Bottom Line: For HIV testing in the ED to be substantially adopted, provider perceptions about barriers must be addressed, including concerns about inadequate time and legal ramifications of testing.

Myers et al. Routine Rapid HIV Screening in Six Community Health Centers Serving Populations at Risk. J Gen Intern Med. 2009. 24:1269–74

Federally qualified community health centers (CHC) are a crucial part of the health-care delivery system for predominantly underserved populations. In late 2006 the National Association of Community Health Centers supported the expansion of point-of-care HIV testing programs across North Carolina, South Carolina, and Mississippi. Myers and colleagues conducted a before-and-after analysis of data from six CHCs who adopted routine rapid HIV screening.[27] Efforts included redesign of patient flow, written protocol development, clinic in-service trainings, on-site technical assistance, follow-up support, and dissemination of patient educational materials.

During the year prior to program rollout, approximately 3% (N=3,078) of patients seen for care were tested for HIV. During the 13 months after rollout, 16,148 were offered testing, and 10,769 (66%) accepted. Thirty-nine persons tested positive on rapid test results, of whom 20 were confirmed HIV-positive and 17 were new HIV diagnoses. Rates of infection were lower than expected by the investigators, but still higher than thresholds disseminated by the CDC for identifying health-care settings eligible for routine testing. Twelve of the 17 newly diagnosed patients were successfully linked to HIV care. Patients who were white were significantly less likely than African-Americans or Latinos to be HIV tested, as were the oldest patients (those aged 55–64 years). While false positives occurred at rates to be expected for the rapid tests, staff were not fully prepared for false positives, and had to develop procedures for understanding and communicating the inherent uncertainty of the initial, preliminary rapid test results.

The Bottom Line: Routine point of care testing markedly increased screening rates in community health centers.

WHAT TO START IN ANTIRETROVIRAL NAÏVE PATIENTS

Table 1 summarizes the current (2009) US Department of Health and Human Services (DHHS) recommendations on which combinations of antiretroviral medications to initiate in the naïve patient.[28]

WHEN TO START ANTIRETROVIRAL THERAPY

Mari Kitahata et al. Effect of Early Versus Deferred Antiretroviral Therapy for HIV on Survival. N Engl J Med. 2009. 360:1815-1826

When to Start Consortium Timing of Initiation of Antiretroviral Therapy in AIDS-Free HIV-1-Infected Patients: A Collaborative Analysis of 18 HIV Cohort Studies. Lancet. 2009. 373:1352–63

When to start antiretroviral therapy (ART) has been an ongoing controversy.[29,30] The decision regarding when to start therapy has to balance anticipated benefits with potential side effects, impact on quality of life, and risk of therapeutic burnout and virologic resistance. Potential benefits of starting early include: modulation of the inflammatory response, lowering of the viral set point, higher rates of immune reconstitution, and use of treatment as prevention.[31,32]

Kitahata, writing for the NA-ACCORD investigators, investigated the effects of starting therapy at different levels of CD4 counts[33]. The NA-ACCORD group consists of 22 clinical cohorts in North America. They identified antiretroviral naïve

Table 1. Antiretroviral Regimens Recommended for Treatment-Naïve Patients. Adapted from Table 5a, Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents, Available at: http://aidsinfo.nih.gov/contentfiles/AdultandAdolescentGL.pdf, Page 39

Preferred Regimens: (Regimens with optimal and durable efficacy, favorable tolerability and toxicity profile, and ease of use)

Non-nucleoside reverse transcriptase inhibitor -based regimen
efavirenz/tenofovir/emtricitabine†
Protease inhibitor-based regimens
atazanavir/ritonavir + tenofovir/emtricitabine†
darunavir/ritonavir (once daily) + tenofovir/emtricitabine†
Integrase strand transfer inhibitor -based regimen
raltegravir + tenofovir/emtricitabine†
Preferred regimen for pregnant women
lopinavir/ritonavir (twice daily) + zidovudine/lamivudine†

Alternative Regimens:
(Regimens that are effective and tolerable but have potential disadvantages compared with preferred regimens. An alternative regimen may be the preferred regimen for some patients.)

Non-nucleoside reverse transcriptase inhibitor -based regimens
efavirenz + (abacavir or zidovudine)/lamivudine†
nevirapine + zidovudine/lamivudine†
Protease inhibitor-based regimens
atazanavir/ritonavir + (abacavir or zidovudine)/lamivudine†
fosamprenavir/ritonavir (once or twice daily) + either [(abacavir or zidovudine)/lamivudine†] or tenofovir/emtricitabine†
lopinavir/ritonavir (once or twice daily) + either [(abacavir or zidovudine)/lamivudine†] or tenofovir/emtricitabine†
saquinavir/ritonavir + tenofovir/emtricitabine†

Acceptable Regimens:
(Regimens that may be selected for some patients but are less satisfactory than preferred or alternative regimens.)

Non-nucleoside reverse transcriptase inhibitor -based regimen
efavirenz + didanosine + (lamivudine or emtricitabine)
Protease inhibitor-based regimen
Atazanavir + (abacavir or zidovudine)/lamivudine†

Comments:
Efavirenz should not be used during the first trimester of pregnancy or in women trying to conceive or not using effective and consistent contraception.
Atazanavir/ritonavir should not be used in patients who require >20 mg omeprazole equivalent per day.
Nevirapine:
- Should not be used in patients with moderate to severe hepatic impairment
- Should not be used in women with pre-ARV CD4 >250 cells/mm^3 or men with pre-ARV CD4 >400 cells/mm^3

Abacavir:
- Should not be used in patients who test positive for HLA-B*5701
- Use with caution in patients with high risk of cardiovascular disease or with pretreatment HIV-RNA >100,000 copies/mL

Once-daily lopinavir/ritonavir is not recommended in pregnant women.

Efavirenz + didanosine + emtricitabine or lamivudine has only been studied in small clinical trials.
Atazanavir/ritonavir is generally preferred over unboosted atazanavir. Unboosted atazanavir may be used when ritonavir boosting is not possible.

†*lamivudine may substitute for emtricitabine or vice versa. Abbreviations: ARV=antiretroviral*

patients who began therapy at CD4 counts greater than 500 cells/mm^3, at CD4 counts between 350 and 500 cells/mm^3, and patients who chose to defer therapy between 1996 and 2006. The main outcome was survival comparing earlier therapy (at either greater than 500 cells/mm^3 or between 350 and 500 cells/mm^3) with deferred therapy. At both levels, deferred treatment was associated with increased death rates [500 cells/mm^3: RR 1.94; 95% CI, 1.37 to 2.79, 350–500 cells/mm^3: RR: 1.69; 95% confidence interval (CI), 1.26 to 2.26].

Evidence supporting starting ART above 350 cells/mm^3 was also recently released by the ART-Cohort Collaboration.[34] They found an increased hazard of death (HR 1.28, 95% CI 1.04–1.57) for starting ART in the 250–350 cells/mm^3 CD4 count range versus 350–450 cells/mm^3. It should be noted that patients with a history of injection drug use were excluded from this study, which may account for the lower hazard ratio. The ART-CC study did not observe a benefit of starting above 450 cells/mm^3, and it is unclear why there is a discrepancy at this higher range. Even larger differences were seen for comparisons with lower CD4 counts at time of treatment initiation.

Because both of these studies are based on observational cohort data, the potential for unmeasured confounders exists. ART initiation is no doubt strongly associated with health status, physician expectations, and patient health seeking behaviors. Both studies adjusted for or excluded patients with a history of injection drug use as a major potential confounder. The NA-ACCORD investigators performed a sensitivity analysis that found a large confounder effect would be needed to reduce or negate their findings. However, an effect of this magnitude was uncovered by the Women's Health Initiative trial of hormone replacement therapy in primary prevention of coronary heart disease compared to prior observational cohort studies.[35] An RCT of ART, the Strategic Timing of AntiRetroviral Treatment (START) trial, is underway and may help to establish the validity of the guidelines.

The Bottom Line: Antiretroviral therapy should be initiated once the CD4 declines below 500 cells/mm^3 and may be beneficial at CD4 counts above 500 cell/mm^3.

Summary and Implications for Practice

Prevention.

- Much work remains to be done, particularly regarding methods initiated and controlled by women, as well as prevention of transmission between serodiscordant couples. Further work is needed to evaluate sexual transmission prevention strategies among men who have sex with men.
- These studies highlight the importance of continuing to recommend condom use as a proven HIV prevention strategy.
- Further research on optimal vaccination regimens, immune responses to vaccination, and the safety and efficacy of vaccinations in specific target populations is still needed.

Testing.

- Routine point of care testing can markedly increase screening rates.
- Attention to issues of stigma and clear communication about results are important for both patients and providers.
- Provider perceptions must be addressed, including concerns about inadequate time and legal ramifications.

Initiation of Treatment. The studies presented both favor earlier treatment in the 350–500 cells/mm^3 range and possibly above. The combined data have led to a change in DHHS Guidelines[28]:

- As before, ART should be initiated if the CD4 count is less than 350 cells/mm^3 or in cases of pregnancy, HIV-

associated nephropathy, and hepatitis B virus (HBV) coinfection when treatment of HBV is indicated.

- Now, the expert panel recommends starting therapy between 350–500 cells/mm^3 and is divided on therapy at CD4 counts above 500 cells/mm^3 with 50% in favor and 50% viewing it as optional.
- It is important to note that most patients are diagnosed with a CD4 under 350 cells/mm^3. Increased adoption of routine testing may identify those with HIV at earlier CD4 counts.

Acknowledgements: *Dr. Sullivan was a Robert Wood Johnson Physician Faculty Scholar during the conduct of this study.*

This update was presented at the meeting of the Society of General Internal Medicine on April 30, 2010.

Conflict of Interest: *Dr. Cofrancesco has been a consultant for Gilead Sciences. No other disclosures.*

Corresponding Author: *Amina A. Chaudhry, MD, MPH; Department of Medicine, Johns Hopkins University School of Medicine, MFL Building, West Tower, Room 528, 5200 Eastern Avenue, Baltimore, MD 21224, USA (e-mail: amina.chaudhry@jhu.edu).*

REFERENCES

1. **Weiss H.** Epidemiology of herpes simplex virus type 2 infection in the developing world. Herpes. 2004;11(SUPPL. 1):24A–35A.
2. **Mosca JD, Bednarik DP, Raj NBK.** Herpes simplex virus type-1 can reactivate transcription of latent human immunodeficiency virus. Nature. 1987;325:67–70.
3. **Kucera LS, Leake E, Iyer N, Raben D, Myrvik QN.** Human immunodeficiency virus type 1 (HIV-1) and herpes simplex virus type 2 (HSV-2) can coinfect and simultaneously replicate in the same human CD4$^+$ cell: effect of coinfection on infectious HSV-2 and HIV-1 replication. AIDS Res Hum Retroviruses. 1990;6:641–647.
4. **Diaz J-J, Duc Dodon M, Schaerer-Uthurralt N, et al.** Post-transcriptional transactivation of human retroviral envelope glycoprotein expression by herpes simplex virus Us11 protein. Nature. 1996;379:273–277.
5. **Heng MCY, Heng SY, Allen SG.** Co-infection and synergy of human immunodeficiency virus-1 and herpes simplex virus-1. Lancet. 1994;343:255–258.
6. **Baeten JM, McClelland RS, Corey L, et al.** Vitamin A supplementation and genital shedding of herpes simplex virus among HIV-1-infected women: a randomized clinical trial. J Infect Dis. 2004;189(8):1466–1471 [cited 1 June 2010].
7. **Schacker T, Ryncarz AJ, Goddard J, Diem K, Shaughnessy M, Corey L.** Frequent recovery of HIV-1 from genital herpes simplex virus lesions in HIV-1-infected men. J Am Med Assoc. 1998;280:61–66.
8. **Schacker T, Zeh J, Hu H, Shaughnessy M, Corey L.** Changes in plasma human immunodeficiency virus type 1 RNA associated with herpes simplex virus reactivation and suppression. J Infect Dis. 2002;186:1718–1725.
9. **Mbopi-Keou F, Gresenguet G, Mayaud P, et al.** Interactions between herpes simplex virus type 2 and human immunodeficiency virus type 1 infection in african women: opportunities for intervention. J Infect Dis. 2000;182:1090–1096.
10. **Gray RH, Wawer MJ, Brookmeyer R, et al.** Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. Lancet. 2001;357:1149–1153.
11. **Delany S, Mlaba N, Clayton T, et al.** Impact of aciclovir on genital and plasma HIV-1 RNA in HSV-2/HIV-1 co-infected women: a randomized placebo-controlled trial in South Africa. AIDS. 2009;23:461–469.
12. **Dunne EF, Whitehead S, Sternberg M, et al.** Suppressive acyclovir therapy reduces HIV cervicovaginal shedding in HIV- and HSV-2-infected women, Chiang Rai, Thailand. J Acquir Immune Defic Syndr. 2008;49:77–83.
13. **Nagot N, Ouédraogo A, Foulongne V, et al.** Reduction of HIV-1 RNA levels with therapy to suppress herpes simplex virus. N Engl J Med. 2007;356:790–799.
14. **Zuckerman RA, Lucchetti A, Whittington WL, et al.** Herpes simplex virus (HSV) suppression with valacyclovir reduces rectal and blood plasma HIV-1 levels in HIV-1/HSV-2-seropositive men: a randomized, double-blind, placebo-controlled crossover trial. J Infect Dis. 2007;196:1500–1508.
15. **Celum C, Wald A, Lingappa JR, et al.** Acyclovir and transmission of HIV-1 from persons infected with HIV-1 and HSV-2. N Engl J Med. 2010;362:427–439.
16. WHO/UNAIDS. New Data on Male Circumcision and HIV Prevention: Policy and Programme Implications. 2007. Available at: http://data.unaids.org/pub/Report/2007/mc_recommendations_en.pdf [cited 1 June 2010].
17. **Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, Puren A.** Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 trial. PLoS Med. 2005;2:1112–1122.
18. **Bailey RC, Moses S, Parker CB, et al.** Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. Lancet. 2007;369:643–656.
19. **Gray RH, Kigozi G, Serwadda D, et al.** Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. Lancet. 2007;369:657–666.
20. **Wawer MJ, Makumbi F, Kigozi G, et al.** Circumcision in HIV-infected men and its effect on HIV transmission to female partners in Rakai, Uganda: a randomised controlled trial. Lancet. 2009;374:229–237.
21. Joint United Nations Programme on HIV/AIDS. UNAIDS fact sheet: Subsaharan Africa. 2007.
22. **Van Damme L, Govinden R, Mirembe FM, et al.** Lack of effectiveness of cellulose sulfate gel for the prevention of vaginal HIV transmission. N Engl J Med. 2008;359:463–472.
23. **Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, et al.** Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N Engl J Med. 2009;361:2209–2220.
24. **Bokhour BG, Solomon JL, Knapp H, Asch SM, Gifford AL.** Barriers and facilitators to routine HIV testing in VA primary care. J Gen Intern Med. 2009;24:1109–1114.
25. **Branson BM, Handsfield HH, Lampe MA, et al.** Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. MMWR.Recommendations and reports: Morbidity and mortality weekly report.Recommendations and reports/Centers for Disease Control. 2006;55:1-17. Available at: http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5514a1.htm. Accessed 3 November 2010.
26. **Arbelaez C, Wright EA, Losina E, et al.** Emergency provider attitudes and barriers to universal HIV testing in the emergency department. J Emerg Med. 2010 (electronic publication ahead of print). Accessed 1 June 2010.
27. **Myers JJ, Modica C, Dufour M-K, Bernstein C, McNamara K.** Routine rapid HIV screening in six community health centers serving populations at risk. J Gen Intern Med. 2009;24:1269–1274.
28. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. 2009 December 1, 2009:1-161. Available at: http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf. Accessed 3 November 2010.
29. **Ho DD.** Time to hit HIV, early and hard. N Engl J Med. 1995;333:450–451.
30. **Henry K.** The case for more cautious, patient-focused antiretroviral therapy. Ann Intern Med. 2000;132:306–311.
31. **Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG.** Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. Lancet. 2009;373:48–57.
32. **Kelley CF, Kitchen CMR, Hunt PW, et al.** Incomplete peripheral CD4+ cell count restoration in HIV-infected patients receiving long-term antiretroviral treatment. Clin Infect Dis. 2009;48:787–794.
33. **Kitahata MM, Gange SJ, Abraham AG, et al.** Effect of early versus deferred antiretroviral therapy for HIV on survival. N Engl J Med. 2009;360:1815–1826.
34. Timing of initiation of antiretroviral therapy in AIDS-free HIV-1-infected patients: a collaborative analysis of 18 HIV cohort studies. The Lancet. 2009;373:1352-63.
35. **Manson JAE, Hsia J, Johnson KC, et al.** Estrogen plus progestin and the risk of coronary heart disease. N Engl J Med. 2003;349:523–534.

Detecting Patient Safety Indicators: How Valid Is "Foreign Body Left During Procedure" in the Veterans Health Administration?

Qi Chen, MPH, Amy K Rosen, PhD, Marisa Cevasco, MD, MPH, Marlena Shin, JD, MPH, Kamal MF Itani, MD, FACS, Ann M Borzecki, MD, MPH

BACKGROUND: The Agency for Healthcare Research and Quality (AHRQ) developed patient safety indicator (PSI) 5, "Foreign body left during procedure," to flag accidental foreign bodies in surgical and medical procedures. This study examined how well this indicator identifies true foreign body events in the Veterans Health Administration (VA).

STUDY DESIGN: This was a retrospective study within 28 selected VA hospitals from fiscal year 2003 to 2007. Trained abstractors reviewed medical charts flagged by PSI 5 and determined true foreign body cases. We calculated the positive predictive value (PPV) of this indicator and performed descriptive analyses of true positive and false positive cases.

RESULTS: Of the 652,093 eligible cases, 93 were flagged by PSI 5 (0.14 per 1,000). Forty-two were true positives, yielding a PPV of 45% (95% CI 35% to 56%). False positives were due to a foreign body that was present on admission (57%) or coding errors (43%). True foreign bodies were associated with surgical (n = 23) and medical (n = 19) procedures. The most common type of surgical foreign body was a sponge (52%). Overall, approximately 40% of foreign bodies were related to a device failure or malfunction (30% surgical vs 53% medical foreign bodies). Postoperative complications included pain (24%), infection (12%), adhesions (5%), and bowel obstruction (5%).

CONCLUSIONS: The reported rate of foreign body events as detected by PSI 5 is low in the VA, but occurs in both surgical and medical procedures. Despite widespread implementation of surgical counts, quality improvement efforts should focus on novel ways to eliminate this "never event" from operations. Future studies are needed to better understand the preventability of medical procedure-associated foreign bodies and particularly, device failure-related foreign bodies. (J Am Coll Surg 2011;212:977–983. © 2011 by the American College of Surgeons)

According to the National Quality Forum (NQF), a "foreign body" unintentionally left in a patient during a procedure is one of the serious reportable events (ie, NQF "never events").[1] It is a rare event, with an estimated incidence rate of 1 in 5,000 operations,[2,3] but has a reported mortality rate as high as 11% to 35%.[4] Risk factors associated with this event include incorrect counts, surgical team transitions, multiple procedures,[5] unplanned changes in the operation, the emergent nature of procedure, and high body mass index.[6] Theoretically, foreign body events can be eliminated by instituting standardized processes of care such as surgical counts of sponges and instruments before and at the end of each operation. However, despite widespread adoption of standardized processes, foreign body events still occur.

Given its potential preventability, the Agency for Healthcare Quality and Research (AHRQ) developed patient safety indicator (PSI) 5, "Foreign body left during procedure," to detect foreign body events using adminis-

Disclosure information: Nothing to disclose.

Supported by Department of Veterans Affairs Health Services Research and Development service grant number SDR 07-002.

A portion of this work was presented at AcademyHealth annual research meeting, Boston, MA, June 2010.

Received December 3, 2010; Revised February 4, 2011; Accepted February 4, 2011.

From the Center for Organization, Leadership, and Management Research (COLMR) (Chen, Rosen, Shin) and the Department of Surgery, VA Boston Healthcare System (Cevasco, Itani); the Department of Health Policy and Management, Boston University School of Public Health (Chen, Rosen, Borzecki); the Department of Surgery, Brigham and Women's Hospital (Cevasco); Boston University School of Medicine (Rosen, Itani, Borzecki); and Department of Surgery, Harvard Medical School (Itani), Boston, MA; and the Center for Health Quality, Outcomes and Economic Research (CHQOER), Bedford VAMC, Bedford, MA (Borzecki).

Correspondence address: Amy K Rosen, PhD, Center for Organization, Leadership and Management Research (COLMR), 150 S Huntington Ave, Boston MA 02130. email: akrosen@bu.edu or amy.rosen2@va.gov

ISSN 1072-7515/11/$36.00
doi:10.1016/j.jamcollsurg.2011.02.003

Abbreviations and Acronyms

AHRQ = Agency for Healthcare Research and Quality
CMS = Centers for Medicare and Medicaid Services
EMR = electronic medical record
IRR = inter-rater reliability
NQF = National Quality Forum
POA = present on admission
PPV = positive predictive value
PSI = patient safety indicator
VA = Veterans Health Administration

trative data. This indicator is defined as discharges with ICD-9-CM codes for foreign body left in during a procedure in any secondary diagnosis field of surgical and medical discharges. ICD-9-CM codes for foreign body used in this PSI include 998.4 (foreign body accidentally left during a procedure not elsewhere classified), 998.7 (acute reaction to foreign substance accidentally left during a procedure not elsewhere classified), and external cause of injury codes (E-codes) E871.x (foreign object left in body during procedure).[7] Similarly, the Centers for Medicare and Medicaid Services (CMS) is currently using an administrative data-based measure (hospital acquired conditions, also known as CMS "never events") to track rates of foreign bodies associated with surgical procedures.[8] However, PSI 5 is the only indicator that flags both surgical and medical foreign bodies.

The accuracy of PSI 5 in identifying true events is unknown. Such information is necessary if this measure is to be used to improve the quality of care. The purpose of this study was to evaluate the positive predictive value (PPV) of this indicator in the Veterans Health Administration (VA) (ie, to evaluate the degree to which the PSI flagged case represents a true foreign body based on data obtained from the medical record). To our knowledge, this is the first study to validate this indicator. We also aimed to examine clinical circumstances and outcomes associated with these true foreign body events.

METHODS

Study design

This was a retrospective observational study using VA inpatient administrative data and electronic medical record (EMR) data from VA fiscal years 2003 to 2007 (from October 1, 2002 to September 30, 2007).[9,10] The required Institutional Review Board approvals from the Bedford VA Medical Center and the VA Boston Healthcare System were obtained for this study.

Study population

We applied the PSI software, version 3.1a, to the VA inpatient dataset to obtain counts of individual PSIs and composite rates. We then selected 28 of 158 VA acute-care hospitals using a stratified sampling method to obtain a diverse sample of VA hospitals. Full details of sampling are provided elsewhere.[9]

PSI 5 technical specifications (v. 3.1)

The denominator of this indicator includes all medical and surgical discharges of patients age 18 or older except for patients with ICD-9-CM codes for foreign body left in during procedure (998.4, 998.7, or E871.x) in the principal diagnosis field or secondary diagnosis present on admission. The numerator includes any discharges with ICD-9-CM codes for foreign body in any secondary diagnosis field among cases meeting the inclusion and exclusion rules for the denominator.[7]

Medical record abstraction

Two trained nurses used standardized abstraction instruments modified from AHRQ-developed instruments to review selected EMRs for the occurrence of a foreign body event and to obtain patient demographics.[9,10] If a foreign body event occurred (ie, a true positive), records were further abstracted for risk factors, clinical circumstances surrounding the procedure associated with the foreign body, and patient outcomes. If no foreign body event occurred (ie, a false positive), abstraction stopped once the reason for exclusion was determined. To assess inter-rater reliability (IRR), both nurses reviewed the same records in groups of 5 until they achieved an average observed agreement of at least 90% across all questions (n = 41) in all records, after which abstraction proceeded independently. IRR was also assessed on 5 charts toward the end of the abstraction process to make sure abstractor reliability did not drift. Further details of the abstraction process are available elsewhere.[9]

Specific to this PSI, we performed 2 early rounds of IRR assessment in order to have enough records to adequately evaluate all questions. Average observed agreement for these 2 rounds combined was 96%. Agreement on late IRR testing was 98%.

Analysis

We calculated PPV as the rate of true positives divided by the total number of flagged cases and derived 95% confidence intervals. We also examined the PPVs of ICD-9-CM codes 998.4, 998.7, and the group of E-codes separately, due to concerns regarding the validity of E-codes in previous.[11,12] We reviewed demographics (age, sex, race and eth-

Table 1. Positive Predictive Values (PPVs) Based on Different Foreign Body ICD-9-CM Codes

ICD-9-CM code(s)	True positives flagged by this code(s), n	False positives flagged by this code(s), n	Positive predictive value, %	95% CI
Overall	42	51	45	35–56
Overall without E871.x	36	37	49	37–61
998.4 only	22	20	52	36–68
998.7 only	1	5	17	0–64
E871.x only	6	14	30	12–54

nicity) in the total sample of 93 patients. Specific to false positives, we examined the type of foreign bodies and the reasons why they were incorrectly flagged by the PSI algorithms. For true positives, we performed descriptive analyses of continuous and categorical variables, including characteristics of foreign bodies, nature of the procedures, risk factors, and patient outcomes, based on whether the foreign body was associated with a surgical or a medical procedure. All statistical analyses were performed using SAS version 9.1 (SAS Institute, Inc).

RESULTS

Across all VA hospitals, a total of 290 foreign body cases out of 2,342,690 eligible discharges were flagged by PSI 5 (0.12 per 1,000); within the 28 selected hospitals, 93 of 652,093 eligible discharges were flagged (0.14 per 1,000). Of the 93 flagged cases, 42 were true positives, yielding a PPV of 45% (95% CI 35% to 56%). ICD-9-CM code 998.4 had a higher PPV (52%) than both ICD-9-CM code 998.7 (17%) and the E-codes (30%), although the confidence intervals overlapped. The overall PPV increased to 49% when we excluded E-codes (Table 1).

Demographics

Flagged patients had an average age of 67 years (SD 11 years) and a median length of stay of 7 days (range 1 to 91 days). Ninety-two patients were male and 73% were white. Demographic characteristics were similar in true positives and false positives (Table 2).

Reasons for false positives

Of the 51 false positives, 29 (57%) were foreign bodies that were present on admission (POA) and 22 (43%) had no documentation of an accidental foreign body event in the EMR during the index hospitalization. Of these 22 cases, 5 false positives represented intentional foreign bodies; 2 (4%) of these had laparotomy pads intentionally left in and 3 had "feeding tubes" (ie, percutaneous endoscopic gastrostomy tubes) inserted. We did not find any evidence of foreign bodies in the other 17 cases. Foreign bodies that were POA included sponges, pads, and fragments of guidewires, drains, and leads— similar to the foreign body types in true positives.

To better understand the accuracy of each individual ICD-9-CM code (998.4, 998.8, and E871.x), we further analyzed the false positive cases flagged by 998.7 or E871.x separately. Of the 5 false positives flagged only by ICD-9-CM code 998.7 ("acute reaction to foreign substance accidentally left during a procedure not elsewhere classified"; Table 1), 2 were POA, and the other 3 were "feeding tubes" that did not cause any postoperative complications. Fourteen false positives were flagged only by external cause-of-injury code E871.x, "foreign object left in body during procedure." (Note that "accidentally or unintentionally" is not incorporated into this definition.) Of these 14 false positives, 10 patients had no evidence of an accidental foreign body, and 1 patient had 3 laparotomy pads intentionally left in an open abdomen to ensure hemostasis during surgery. (Note that the other patient who had laparotomy

Table 2. Patient Demographics (n = 93)

Variable	All flagged cases (n = 93)	True positives (n = 42, 45%)	False positives (n = 51, 55%)
Age, mean (SD), y	67 (11)	67 (12)	67 (11)
Length of stay, median (range), d	7 (1–91)	7 (1–80)	8 (1–91)
Male sex, n (%)	92 (99)	41 (98)	51 (100)
Race, n (%)			
White	68 (73)	31 (74)	37 (73)
Black	13 (14)	5 (12)	8 (16)
Hispanic	4 (4)	2 (5)	2 (4)
Other/missing	8 (9)	4 (9)	4 (8)

pads intentionally left in was flagged by ICD-9-CM code 998.4.)

True positive analyses

Of the 42 true positives, 23 foreign bodies (55%) were related to a surgical procedure; 15 (61%) of these were related to an abdominopelvic procedure. Twenty-one surgical foreign bodies were associated with the original operation, and the other 2 foreign bodies (drain fragments) occurred during postoperative drain removal. Sponges were the most common type of foreign body left behind during surgical procedures (n = 12, 52%); the remainder were instrument or device fragments (Foley tips, drill tips, a steel fragment from a resectoscope, and drain fragments). All surgical procedures had documentation of correct surgical counts (sponge, instrument, and sharp counts) except 4, which documented a disagreement in counts (1 in the final sponge count and 3 in the final instrument count). Intraoperative radiologic survey was performed in these 4 discrepant count cases, and in 3 other cases based on surgeon's concern. Overall, 9 foreign bodies (39%) were discovered at the time of procedure (7 by intraoperative radiologic survey, 1 by surgeon's manual exploration of surgical site, and 1 drain fragment that was discovered immediately after drain removal). Of the 8 foreign bodies that were detected during the original operation, 4 were discovered before skin closure and 4 afterwards. However, only 2 foreign bodies were removed before the patient left the operating room; the rest (n = 6) were left in the patient and scheduled for removal in subsequent procedures based on the surgeon's decision (eg, the patient was hemodynamically unstable and required additional stabilization before object retrieval). Among the other 14 true positives (61%) who had foreign bodies discovered postoperatively, 7 were discovered during investigation of symptoms, 6 were discovered during routine postoperative screening in patients with no symptoms, and 1 was discovered incidentally during a subsequent operation. Seven of the surgical foreign bodies (30%) were related to device failure or malfunction (an instrument broke during the procedure or a device fragment was accidentally left in patient) (Table 3).

The remaining true positives (n = 19) were related to a medical procedure, including cardiac catheterization (n = 12), central line placement (n = 3), and pacemaker placement (n = 2). The most common types of foreign bodies were guidewires or guidewire fragments (n = 13, 68%). Ten of these (53%) were related to device failure or malfunction. Eleven foreign bodies (58%) were discovered at the time of procedure (the physician found part of the instrument was missing when he or she pulled it out) (Table 4).

Patient symptoms related to retained foreign bodies included pain (n = 10), infection (n = 5), adhesions (n = 2), and bowel obstruction (n = 2) (Table 5). Two patients died during the index hospitalization, but neither death was due to the retained foreign body.

Table 3. Characteristics of Foreign Body and Procedure in Surgical Procedure Related True Positives (n = 23)

Variable	n	%
Type of surgical foreign body		
Sponge or gauze	12	52
Instrument or device fragments (Foley tips, drill tips, and a steel fragment from a resectoscope)	7	30
Drain fragments	2	9
Other*	2	9
How foreign body was discovered		
At the time of the original procedure	9	39
Routine postprocedure screen or test without presenting symptoms	7	30
Clinical detection with presenting symptoms	6	26
During subsequent surgery related to signs/symptoms/complications	2	9
During subsequent drain removal procedure	2	9
Incidental discovery during subsequent surgery	1	4
Surgical site		
Abdomen and pelvis	14	61
Spine and extremity (1 vascular and 5 orthopaedic)	6	26
Chest	3	13
Intraoperative radiographic study performed†	7	30
Operative site reopened to look for foreign body or remove foreign material:		
Before leaving operating room and after skin closed	2	9
After leaving operating room	12	52
Unable to determine	1	4
Obesity (body mass index ≥30 kg/m^2)	8	35
Multiple surgical teams involved in procedure	3	13
Procedure performed on emergent or urgent basis	2	9
Unplanned change in procedure‡	3	13
Procedure performed on weekend or weekday night	4	17
Foreign body related device failure or malfunction	7	30

*1) The basket used to retrieve kidney stone; 2) A piece of tape from camera drape during esophagectomy.
†Three radiology studies were performed before skin closure, 4 were after.
‡1) Video-assisted thoracic surgery converted to thoracotomy; 2) unplanned splenectomy; 3) laparoscopic converted to open cholecystectomy.

DISCUSSION

"Unintentionally retained foreign body" was a rare event in the VA. The reported rate of this safety event by PSI 5 was 0.14 per 1,000 cases in the 28 sample hospitals and 0.12 per 1,000 cases across all VA hospitals. As the first study to examine the accuracy of the AHRQ PSI 5, "foreign body left during procedure," we found that the PPV (45%) of this indicator was relatively low compared with other examined PSIs.[9,10]

Table 4. Characteristics of Foreign Body and Procedure in Medical Procedure Related True Positives (n = 19)

Variable	n	%
Type of medical foreign body		
Guidewire or guidewire fragment	13	68
Other instrument fragments	5	26
Stent	1	5
Procedure type		
Cardiac catheterization	12	63
Pacemaker placement or removal	3	16
Central line placement	2	11
Other (eg, replace gastrostomy tube)	2	11
Rank of person performing this procedure*		
Attending physician	2	11
Physician-in-training	2	11
Physician, unknown ranking	5	26
Not documented	10	53
How foreign body was discovered		
At the time of procedure	11	58
Routine postoperative physical examination or radiology without presenting symptoms	6	32
Clinical detection with symptoms	2	11
Procedure performed on emergent or urgent basis	3	16
Procedure was performed on weekend or weekday night	3	16
Foreign body related to device failure or malfunction	10	53

*We collected only the ranking of physicians in medical procedures.

The reasons for false positives were due to conditions being POA (57%) and coding errors (43%). Such a high rate of POA conditions highlights the need to implement POA codes in the VA; the frequency of coding errors underscores existing concerns about coding accuracy, particularly the validity of E-codes.[11,12] Presumably because of such concerns, only codes 998.4 and 998.7 are used for the CMS hospital-acquired condition measure, "foreign object obtained after surgery." This measure also assumes that POA coding is in place and being implemented correctly.[8] Assuming appropriate POA coding and using only the 998.x codes, the PPV of PSI 5 in our study would have increased to 80%, a more moderate PPV.

Our detailed review of false positive cases raised some additional concerns about coding. First, it is unclear whether VA coders are familiar with how to use these foreign body codes. For example, we found that 1 patient had 3 laparotomy pads intentionally left in the open abdomen to ensure hemostasis, but it was coded as an accidental foreign body event. It might be difficult for a coder to determine whether a foreign body was intentionally or unintentionally left in a patient if he or she did not have a strong clinical and/or surgical background or the medical record did not provide explicit details about the scenario. Second, implementation of POA codes in the VA would undoubtedly improve the validity of PSI 5 for detecting true in-hospital events. However, one might argue that from the perspective of measuring performance, it is also important to capture foreign bodies that are POA, because they might reflect true complications of care that occurred during a previous admission or in the outpatient setting. For example, in our false positive review, we found a patient who underwent low anterior resection of the rectosigmoid at Hospital A. After discharge he developed low-grade fevers and left upper quadrant abdominal pain. On return to the hospital 1 month later for a follow-up examination, an abdominopelvic CT showed a retained sponge in his left upper quadrant. He was admitted and brought to the operating room to remove this foreign body. After implementing POA codes in the VA, this case would not be flagged as a foreign body event. However, it is a true safety event that occurred in Hospital A, and should be counted when we evaluate the performance of this facility.

In the true positive review, despite seemingly universal compliance with manual counting protocols for sponges in VA, we still found 12 cases of sponges left in a patient during an operating room procedure when pre- and post-sponge counts supposedly "agreed." This highlights the issue of "false correct counts" (final counts erroneously thought to be correct), suggesting human error in surgical counts. Previous studies have also found other types of human errors in surgical counts.[5,6,13,14] For example, Greenberg and colleagues[14] found that 1 in 8 operations involved at least 1 counting discrepancy (defined as "any instance in which a subsequent count does not agree with a previous one") while they were observing procedures in the operating room, and 41% of these discrepancies were due to miscounting (disagreements between pre- and post-counts that are due to an error in counting). Study authors

Table 5. Patient Outcomes in Surgical and Medical Foreign Bodies (n = 42)

	Surgical foreign bodies (n = 23)		Medical foreign bodies (n = 19)	
Patient outcomes	n	%	n	%
Additional pain or discomfort	5	22	5	26
Sepsis or infection, inflammatory process or other acute reaction	4	17	1	5
Adhesions	2	9	0	0
Bowel obstruction	2	9	0	0
Other*	2	9	0	0
No discomfort	15	65	13	68
Additional procedure or surgery to remove foreign body	22	96	19	100

*1) Mental status changes; 2) Pressure ulcer.

suggested that technological solutions may be needed to reduce such surgical count errors. They subsequently conducted a clinical trial with cost-effectiveness analysis, showing that bar-coding of sponges can improve the detection of miscounted and misplaced sponges in operations at an acceptable cost.[15,16] Although more research needs to be done, technological solutions may be one way to reduce foreign body events, at least in surgical settings.

Although surgical foreign bodies, such as sponges left behind, may be eliminated either by standardized counting or introduction of some of the available technologies to track sponges, we found that almost 40% of all foreign bodies (both surgical and medical foreign bodies) were related to a device failure or malfunction. For example, one patient had a Foley catheter placed before arthroscopic shoulder surgery. The Foley balloon broke and a missing portion (about 0.25 × 0.25 inch) was accidentally left in the patient. Device-related foreign body events occurred more often in medical (n = 10) than surgical (n = 7) procedures. Although previous studies have shown the general public health burden of adverse events associated with medical devices,[17-19] none of these have focused on foreign body events. This study was the first to examine device-associated foreign bodies; our results underscore a concern about the preventability of this type of foreign body event. Although operator inexperience may be one of the reasons for device failure or malfunction, it is hard to determine if a sheared-off guidewire occurred due to physician error or other factors. It will be difficult to prevent such device-associated foreign body events if we cannot clearly identify the reason for device failures.

This was also the first study to examine foreign body events related to medical procedures. Although quality improvement efforts should focus on both medical and surgical procedures, further research is necessary to determine the degree to which these medical foreign body events (and as noted, those associated with device failure) are preventable. Until such information is available, any public reporting of foreign body events, as CMS is doing, should be limited to events associated with surgical procedures.[8]

Notably, the timing of the foreign body event is one of the key components in defining a true event. In this study, 2 foreign bodies discovered after surgical site closure but before leaving the operating room were deemed to be true positives based on the AHRQ definition. Surgical studies usually do not recognize this type of scenario as a true event, because the patient was still in the operating room and therefore the procedure was not officially completed.[6,13,14,20] These 2 cases highlight the discrepancy between the surgical and AHRQ definitions of foreign body. Further discussion and clarification may be needed to ensure that the definition of a foreign body event is consistent across coders, surgeons, and health service researchers, especially in comparing provider performance.

As with any study, there are some inherent limitations. We do not report the sensitivity, specificity, or negative predictive value of this PSI because we did not examine patients who were not flagged by PSI 5. Furthermore, all the data used in this study were abstracted from the EMR, where providers may not completely document all relevant clinical details related to the hospitalization. However, as discussed earlier, the strengths of this study are that it is the first to examine the PPV of this PSI, examine foreign body events occurring outside the operating room, and report on foreign body events associated with device failures. There are study design strengths as well: we selected a nationwide representative sample of VA hospitals; we performed explicit review of the EMR to collect data; high abstractor agreement was obtained in IRR tests; false positives were reviewed to identify ways to improve the validity of this PSI; and we examined true positives in detail to try to determine the preventability of this safety event.

CONCLUSIONS

Based on the reported rate of PSI 5, a foreign body unintentionally left in a patient is a rare event in the VA. However, it occurs in both surgical and medical procedures. Because both NQF and CMS have defined a surgical foreign body as a "never event," quality improvement efforts should focus on novel ways of eliminating unintentionally retained foreign bodies. Although medical procedure-associated foreign bodies are also considered as "never events" by NQF, future studies are needed to better understand the preventability of medical foreign bodies, and particularly device failure-related foreign bodies.

Author Contributions

Study conception and design: Chen, Rosen, Cevasco, Shin, Itani, Borzecki

Acquisition of data: Chen, Borzecki

Analysis and interpretation of data: Chen, Cevasco, Itani, Borzecki

Drafting of manuscript: Chen, Borzecki

Critical revision: Chen, Rosen, Cevasco, Shin, Itani, Borzecki

Acknowledgment: Special thanks to: Dijon Fasoli, RN, PhD for conducting inter-rater reliability assessments; Kathleen Hickson, RN, MA for review of the accuracy of the electronic form; Haytham Kaafarani, MD, MPH for assistance with abstraction instrument and guidelines modification and case review; Julie Lynch, MBA, RN for her assistance with data

abstraction and data cleaning; Susan Loveland, MAT for programming expertise and assistance with identifying the hospital sample and flagged cases; Hillary Mull, MPP for technical assistance in development of the electronic abstraction form; and Kristine Ruggiero, RN for her assistance with data abstraction.

REFERENCES

1. National Quality Forum. Serious Reportable Events in Healthcare – 2006 Update. Available at: http://www.qualityforum.org/Publications/2007/03/Serious_Reportable_Events_in_Healthcare-2006_Update.aspx. Accessed February 23, 2011.
2. Cima RR, Kollengode A, Garnatz J, et al. Incidence and characteristics of potential and actual retained foreign object events in surgical patients. J Am Coll Surg 2008;207:80–87.
3. Romano PS, Geppert JJ, Davies S, et al. A national profile of patient safety in U.S. hospitals. Health Aff (Millwood) 2003;22:154–166.
4. Lauwers PR, Van Hee RH. Intraperitoneal gossypibomas: the need to count sponges. World J Surg 2000;24:521–527.
5. Lincourt AE, Harrell A, Cristiano J, et al. Retained foreign bodies after surgery. J Surg Res 2007;138:170–174.
6. Gawande AA, Studdert DM, Orav EJ, et al. Risk factors for retained instruments and sponges after surgery. N Engl J Med 2003;348:229–235.
7. AHRQ. PSI 5 Foreign body left during procedure technical specifications. Agency for Healthcare Research and Quality; 2010. Available at: http://www.qualityindicators.ahrq.gov/downloads/psi/TechSpecs42/PSI%2005%20Foreign%20Body%20Left%20during%20Procedure.pdf. Accessed October 25, 2010.
8. CMS. Hospital-acquired conditions. 2010; Available at: https://www.cms.gov/HospitalAcqCond/06_Hospital-Acquired_Conditions.asp. Accessed February 23, 2011.
9. Borzecki AM, Kaafarani HMA, Utter GH, et al. How valid is the AHRQ patient safety indicator "postoperative respiratory failure"? J Am Coll Surg 2011;212:935–945.
10. Kaafarani HMA, Borzecki AM, Itani KMF, et al. Validity of selected patient safety indicators: opportunities and concerns. J Am Coll Surg 2011;212:924–934.
11. Smith GS, Langley JD. Drowning surveillance: how well do E codes identify submersion fatalities. Inj Prev 1998;4:135–139.
12. Oklahoma State Department of Health. Importance, accuracy, and completeness of external cause of injury coding (E coding) in a sample of Oklahoma traumatic brain injury hospitalizations. Oklahoma State Department of Health, 2009. Available online at: http://www.ok.gov/health/documents/E-coding.pdf. Accessed March 15, 2011.
13. Gawande AA, Zinner MJ, Studdert DM, Brennan TA. Analysis of errors reported by surgeons at three teaching hospitals. Surgery 2003;133:614–621.
14. Greenberg CC, Regenbogen SE, Lipsitz SR, et al. The frequency and significance of discrepancies in the surgical count. Ann Surg 2008;248:337–341.
15. Greenberg CC, Diaz-Flores R, Lipsitz SR, et al. Bar-coding surgical sponges to improve safety: a randomized controlled trial. Ann Surg 2008;247:612–616.
16. Regenbogen SE, Greenberg CC, Resch SC, et al. Prevention of retained surgical sponges: a decision-analytic model predicting relative cost-effectiveness. Surgery 2009;145:527–535.
17. Hefflin BJ, Gross TP, Schroeder TJ. Estimates of medical device--associated adverse events from emergency departments. Am J Prev Med 2004;27:246–253.
18. Samore MH, Evans RS, Lassen A, et al. Surveillance of medical device-related hazards and adverse events in hospitalized patients. JAMA 2004;291:325–334.
19. Wang C, Hefflin B, Cope JU, et al. Emergency department visits for medical device-associated adverse events among children. Pediatrics 2010;126:247–259.
20. Greenberg CC, Gawande AA. Retained foreign bodies. Adv Surg 2008;42:183–191.

Circulating Testosterone and SHBG Concentrations Are Heritable in Women: The Framingham Heart Study

A. D. Coviello, W. V. Zhuang, K. L. Lunetta, S. Bhasin, J. Ulloor, A. Zhang, D. Karasik, D. P. Kiel, R. S. Vasan, and J. M. Murabito

Sections of Preventive Medicine and Epidemiology (A.D.C., R.S.V.), Endocrinology, Diabetes, and Nutrition (A.D.C., S.B., J.U., A.Z.), and General Internal Medicine (J.M.M.), Department of Medicine, Boston University School of Medicine, and Department of Biostatistics (W.V.Z., K.L.L.), Boston University School of Public Health, Boston, Massachusetts 02118; Institute for Aging Research (D.K., D.P.K.), Hebrew SeniorLife, Harvard Medical School, Boston, Massachusetts 02131; and Framingham Heart Study (K.L.L., R.S.V., J.M.M.), Framingham, Massachusetts 01702

Context: Many factors influence the concentration of circulating testosterone and its primary binding protein, SHBG. However, little is known about the genetic contribution to their circulating concentrations in women, and their heritability in women is not well established.

Objective: Our objective was to estimate the heritability of circulating total testosterone (TT), free testosterone (FT), and SHBG in women in families from the Framingham Heart Study.

Methods: Women in the Framingham Heart Study who were not pregnant, had not undergone bilateral oophorectomy, and were not using exogenous hormones were eligible for this investigation. TT was measured using liquid chromatography tandem mass spectrometry and SHBG using an immunofluorometric assay (Delfia-Wallac), and FT was calculated. Heritability estimates were calculated using variance-components methods in Sequential Oligogenic Linkage Analysis Routines (SOLAR) and were adjusted for age, age^2, body mass index (BMI), BMI^2, diabetes, smoking, and menopausal status. Bivariate analyses were done to assess genetic correlation between TT, FT, and SHBG.

Results: A total of 2685 women were studied including 868 sister pairs and 688 mother-daughter pairs. Multivariable adjusted heritability estimates were 0.26 ± 0.05 for FT, 0.26 ± 0.05 for TT, and 0.56 ± 0.05 for SHBG ($P < 1.0 \times 10^{-7}$ for all). TT was genetically correlated with SHBG [genetic correlation coefficient (ρG) $= 0.31 \pm 0.10$] and FT ($\rho G = 0.54 \pm 0.09$), whereas SHBG was inversely correlated with FT ($\rho G = -0.60 \pm 0.08$).

Conclusion: Circulating TT, FT, and SHBG concentrations in women are significantly heritable, underscoring the importance of further work to identify the specific genes that contribute significantly to variation in sex steroid concentrations in women. The strong shared genetic component among pairs of TT, FT, and SHBG concentrations suggests potential pleiotropic effects for some of the underlying genes. **(*J Clin Endocrinol Metab* 96: E1491–E1495, 2011)**

Sex steroids are essential for normal sexual differentiation and reproductive health across the lifespan. Sex steroids influence many age-related chronic diseases in women including osteoporosis (1), metabolic syndrome (MetS) (2, 3), type 2 diabetes (T2DM) (2, 4, 5), and cardiovascular disease (CVD) (6, 7). The relationship between sex steroids and chronic disease is affected by changes in hormone profiles in women, particularly with menopause.

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in U.S.A.

doi: 10.1210/jc.2011-0050 Received January 10, 2011. Accepted June 17, 2011.
First Published Online July 13, 2011

Abbreviations: BMI, Body mass index; CVD, cardiovascular disease; FHS, Framingham Heart Study; MetS, metabolic syndrome; T2DM, type 2 diabetes.

Testosterone and its primary binding protein, SHBG, mediate sex-hormone-sensitive phenotypes in women. Higher circulating testosterone has been associated with incident T2DM (4) and CVD (8). Lower SHBG has been associated with MetS and T2DM (4, 9, 10), whereas higher SHBG has been associated with greater risk for hip fracture (1).

Although circulating testosterone and SHBG are affected by many environmental factors including obesity, smoking, insulin resistance, and T2DM (11), they are believed to be at least partially heritable. In men, heritability estimates range from 25–75% for testosterone and 30–50% for SHBG (12–14). In contrast, little is known about the genetic influences of circulating sex hormones in women. One study estimated the heritability of total testosterone at 39% and SHBG at 56% in postmenopausal sisters and twins (15). The purpose of this investigation was to estimate the heritability of circulating total and free testosterone and SHBG in adult women in families from the Framingham Heart Study (FHS), a multigenerational population-based study.

Subjects and Methods

Study population

The FHS design and participants have been described in detail previously (16). In brief, the FHS recruited a population-based sample of men and women residing in Framingham, MA, in 1948 (n = 5209) with the purpose of prospectively studying CVD. In 1971, children of the original cohort and their spouses were recruited as a second generation (the Offspring Study). In 2002, children of the Offspring cohort were recruited as a third generation. Women from Generation 2 and Generation 3 aged 19 and above with sex steroid hormone measurements were eligible for this study. Women taking estrogens, progestins, and androgens or who were pregnant or had undergone bilateral oophorectomy were excluded. All participants signed informed consent before participation in the FHS. The study was approved by the institutional review board at Boston University Medical Center.

All blood samples were collected in the morning, usually between 0730 and 0930 h after participants fasted approximately 10 h overnight. Samples were aliquoted, frozen, and stored at −80 C. Samples used for testosterone and SHBG measurements were not thawed previously.

Assay measurements

Total testosterone was measured by liquid chromatography tandem mass-spectrometry (17). Mass spectrometry was performed using TSQ Thermo-Finnigan Quantum Ultra (Thermo Fisher Scientific, Waltham, MA). The functional limit of detection, defined as the lowest concentration detected with less than 20% coefficient of variation, was 2 ng/dl. Cross-reactivity with other steroids including dehydroepiandrosterone/dehydroepiandrosterone sulfate, androstenedione, and estradiol was negligible. SHBG was measured with a two-site directed immunofluorometric assay with sensitivity of 0.5 nM and less than 0.1% cross-reactivity with other circulating proteins (Delfia-Wallac, Inc., Turku, Finland). Free testosterone was calculated from total testosterone and SHBG using the law of mass action equations (18).

Statistical analysis

Serum hormones were transformed by rank normalization to minimize skew. Covariates considered in the analysis included age, age^2, body mass index (BMI), BMI^2, T2DM, smoking, and menopausal status. Covariates were defined as follows: T2DM, fasting blood sugar of at least 126 mg/dl or use of diabetes medications; current smoking, yes/no; and postmenopause, at least 1 yr without menses. SAS version 9.1 (SAS Inc., Cary, NC) was used. SOLAR (Sequential Oligogenic Linkage Analysis Routines) statistical software was used to estimate heritability (19). All family relationships including first-degree relatives (mother-daughter, sister-sister pairs) and extended family relationships (aunt-niece, cousins) were analyzed. Heritability estimates were adjusted for the same covariates across all models. Heritability estimates for free and total testosterone were also adjusted for SHBG to assess for genetic influences independent of SHBG. Bivariate analyses were used to examine the genetic correlation between circulating testosterone and SHBG and were adjusted for the above covariates.

Results

The 2685 women studied included first-degree relatives (868 sister-sister, 688 mother-daughter pairs) and extended family relationships (Table 1). The older, predominantly postmenopausal Generation 2 women had higher BMI (28 ± 6 *vs.* 26 ± 6 kg/m^2) and more T2DM (10 *vs.* 2%) than the younger, predominantly premenopausal Generation 3. SHBG was higher in Generation 3, whereas free testosterone was slightly higher in Generation 2.

Heritability estimates (Tables 2 and 3)

Total and free testosterone as well as SHBG showed strong heritability after adjusting for age, BMI, T2DM, current smoking, and menopausal status (Table 2). Both total and free testosterone showed heritability estimates of 0.26 (SE = 0.05). SHBG heritability was estimated at 0.56

TABLE 1. Characteristics of women from the Offspring and Generation 3 cohorts of the FHS

	Offspring Generation 2	Generation 3
Sample (n)	1071	1614
Age (yr)	62 ± 10	41 ± 8
BMI (kg/m^2)	28 ± 6	26 ± 6
Postmenopause (%)	80	10
Smoker (%)	12	16
Diabetes (%)	10	2
Total testosterone (ng/dl)	32.1 ± 21.8	27.9 ± 15.3
Free testosterone (pg/ml)	3.6 ± 2.4	2.9 ± 1.9
SHBG (nmol/liter)	74.2 ± 38.8	87.2 ± 46.6

TABLE 2. Heritability of total and free testosterone and SHBG in adult women from the FHS Generations 2 and 3: univariate analysis

	Total testosterone	SHBG	Free testosterone
n	2671	2677	2671
h^2 (SE)	0.263 (0.053)	0.556 (0.052)	0.259 (0.052)
P value	1.00×10^{-7}	7.29×10^{-30}	3.25×10^{-8}
SHBG-adjusted analysis[a]			
n	2671		2671
h^2 (SE)	0.246 (0.053)		0.203 (0.052)
P value	3.00×10^{-7}		1.24×10^{-5}

Covariates included age, age^2, BMI, BMI^2, menopausal status, diabetes, and current smoking. h^2, Heritability estimate.

[a] Total and free testosterone heritability estimates adjusted for circulating SHBG in addition to age, age^2, BMI, BMI^2, menopausal status, diabetes, and current smoking.

(SE = 0.05). Heritability estimates of circulating free and total testosterone remained significant after adjustment for SHBG: free testosterone heritability was 0.20 (SE = 0.05) and total testosterone heritability was 0.25 (SE = 0.05).

Bivariate analyses (Table 3) showed significant genetic correlation between total testosterone and SHBG [genetic correlation coefficient (ρG) = 0.31; SE = 0.10; $P = 3.32 \times 10^{-3}$] after adjustment for age, BMI, T2DM, current smoking, and menopausal status. Total and free testosterone were also highly genetically correlated (ρG = 0.54; SE = 0.09; $P = 1.69 \times 10^{-3}$) as would be expected because free testosterone, a small fraction of total testosterone concentrations, is calculated from total testosterone and SHBG. SHBG was inversely genetically correlated with free testosterone (ρG = −0.60; $P = 1.93 \times 10^{-8}$), suggesting that specific genes may have opposite effects on their circulating concentrations. Estimates of environmental influences on circulating SHBG and free testosterone were inversely correlated as well (environmental correlation, ρE = −0.37; SE = 0.05; $P = 2.23 \times 10^{-8}$), suggesting that environmental influences may affect SHBG and free testosterone in opposite directions.

The derived phenotypic correlation, ρP, between SHBG and free testosterone is −0.44 (Table 3), which is consistent with the correlation of the two hormones circulating concentrations (Pearson correlation = −0.41, $P < 0.0001$).

TABLE 3. Heritability of total and free testosterone and SHBG in adult women from the FHS Generations 2 and 3: bivariate analysis

	Total testosterone and SHBG	Total testosterone and free testosterone	SHBG and free testosterone
n	2677	2671	2677
ρG (SE)	0.31 (0.10)	0.54 (0.09)	−0.60 (0.08)
P value			
Ho: ρG = 0	3.32×10^{-3}	1.69×10^{-3}	1.93×10^{-8}
Ho: ρG = 1	3.00×10^{-7}	1.54×10^{-9}	1.29×10^{-5}
ρE (SE)	0.10 (0.06)	0.85 (0.017)	−0.37 (0.053)
P value			
Ho: ρE = 0	0.14	2.23×10^{-8}	9.78×10^{-44}
ρP	0.171	0.769	−0.440

Covariates included age, age^2, BMI, BMI^2, menopausal status, diabetes, and current smoking. ρE, Environmental correlation; ρG, genetic correlation; ρP = phenotypic correlation; Ho, null hypothesis.

Discussion

We have shown that total and free testosterone and SHBG are moderately to highly heritable (26–56%) in white women of European ancestry in the FHS. Furthermore, circulating total testosterone and SHBG are genetically correlated, suggesting that the two are influenced by common genes or that genetic influences on SHBG may indirectly influence total testosterone given that SHBG is testosterone's primary binding protein. Free testosterone and SHBG are inversely genetically correlated, suggesting that genes may influence them in opposite directions, which is consistent with the inverse correlation of their circulating concentrations. Given the importance of testosterone and SHBG for both reproductive and nonreproductive health in women, further analyses of the genetic influences of circulating testosterone and SHBG are needed to elucidate risk factors for the diseases influenced by these hormones, particularly osteoporosis, MetS, T2DM, and CVD.

Women experience a rapid decline in estrogen production at the time of menopause. In contrast, testosterone declines at a relatively stable rate across the menopause (20). The menopausal transition is associated with changes in metabolism and body fat distribution consistent with the MetS. Furthermore, there is an increase in osteoporosis (1) and CVD (21) after menopause traditionally thought to be due to estrogen deficiency. However, estrogen and estrogen/progestin treatment in clinical trials have not shown a benefit in CVD risk or mortality and have suggested possible increased risk for CVD and stroke (7). The role of testosterone independent of estrogens in osteoporosis, CVD, and metabolic disorders is not clear. Some studies (including the FHS) (22) have shown that higher testosterone levels are associated with increased prevalence of CVD and T2DM (2, 4). Prospective studies have found higher testosterone to be associated with increased risk of T2DM (4) and CVD (8). Postmenopausal women in the WISE (Women's Ischemia Syndrome Evaluation) study being evaluated for ischemia with a history

of irregular menses and elevated free testosterone had a higher burden of angiographic coronary artery disease and were more likely to have a myocardial infarction or cardiovascular-related death (23). There are very limited data on the genetic influences of circulating testosterone in otherwise healthy women.

SHBG tightly binds testosterone in women. SHBG production and clearance are affected by many factors including age, adiposity, smoking, hormone use, liver disease, insulin resistance, and T2DM, resulting in variable circulating SHBG levels (11). SHBG was thought to function only as a carrier protein for sex hormones, but it has been proposed that SHBG binds to its own receptor and exerts biological effects as well (24, 25). Lower SHBG has recently been shown to predict incident T2DM in women (9–10). SHBG concentrations are higher in premenopausal women compared with postmenopausal women due to the stimulation of SHBG production by the liver in response to higher estrogen levels. Thus, the lower SHBG levels coincident with lower estrogen levels after menopause may be related to the greater metabolic and cardiovascular risk observed after the menopausal transition.

Our adjusted heritability estimate for SHBG of 56% suggests a significant genetic component and is in the range estimated in the San Antonio Heart Study in men and women and extends the observation in the Australian study of postmenopausal sisters to premenopausal women as well (12, 15). Additionally, SHBG had a strong inverse genetic correlation with free testosterone, suggesting that the two share common genes that influence their circulating levels. The candidate gene approach has been used to identify regions of interest that may influence circulating hormone levels in conjunction with disease states. A *SHBG* single-nucleotide polymorphism, rs179994, has been associated with T2DM (odds ratio = 0.94; 95% confidence interval = 0.91–0.97; $P = 2 \times 10^{-5}$) (10). Further study of the genetic factors influencing circulating SHBG levels are warranted to determine whether these genes influence susceptibility to chronic diseases, particularly T2DM and CVD.

The use of mass spectrometry, a state of the art method for measuring testosterone in the low concentrations prevalent in women, is a significant strength of this study. The multigenerational design of the FHS is also a significant strength. Our findings are based on the FHS population, which is Caucasian of European ancestry. Future studies of the genetic influences of circulating sex steroid profiles in women should be conducted in populations of varied racial and ethnic backgrounds.

Summary

Circulating testosterone and SHBG levels are highly heritable in women of white, European ancestry, suggesting strong genetic influences. Further study to elucidate the genetic loci that contribute to the determination of circulating hormone profiles is important given that circulating sex hormone profiles are associated with increased risk for significant chronic disease in women including osteoporosis, T2DM, and CVD. The candidate gene approach is limited by the assumptions underlying the choice of the genes to be analyzed. Genome-wide association scans may be a more fruitful approach for identifying genetic regions that influence sex hormone and SHBG levels that are not encumbered by the bias of *a priori* assumptions about which genes will be important in mediating their synthesis or action. There is evidence of potential pleiotropic genetic effects with single genes affecting both testosterone and SHBG. Understanding pleiotropic genetic effects on circulating hormone concentrations may have important implications for the development of hormone-targeted therapy for chronic diseases.

Acknowledgments

Address all correspondence and requests for reprints to: Andrea Coviello, M.D., M.Sc., Assistant Professor of Medicine, Section of Preventive Medicine and Epidemiology, Section of Endocrinology and Metabolism, Boston University School of Medicine, 760 Harrison Avenue, Harrison Court B91, Boston, Massachusetts 02118. E-mail: coviello@bu.edu.

This work was supported by National Institutes of Health Grants RO1 HL094755 (to A.D.C., S.B., and R.S.V.), RO1 AG31206 (to R.S.V. and S.B.), R21 AG032598 (to W.V.Z., J.M.M., K.L.L., D.P.K., and D.K.), and RO1 AR/AG41398 (to D.P.K.), and by a Foundation Grant from the Centers for Disease Control (to S.B.).

Disclosure Summary: The authors have nothing to disclose.

References

1. **Lee JS, LaCroix AZ, Wu L, Cauley JA, Jackson RD, Kooperberg C, Leboff MS, Robbins J, Lewis CE, Bauer DC, Cummings SR** 2008 Associations of serum sex hormone-binding globulin and sex hormone concentrations with hip fracture risk in postmenopausal women. J Clin Endocrinol Metab 93:1796–1803
2. **Patel SM, Ratcliffe SJ, Reilly MP, Weinstein R, Bhasin S, Blackman MR, Cauley JA, Sutton-Tyrrell K, Robbins J, Fried LP, Cappola AR** 2009 Higher serum testosterone concentration in older women is associated with insulin resistance, metabolic syndrome, and cardiovascular disease. J Clin Endocrinol Metab 94:4776–4784
3. **Shakir YA, Samsioe G, Nyberg P, Lidfeldt J, Nerbrand C, Agardh CD** 2007 Do sex hormones influence features of the metabolic syndrome in middle-aged women? A population-based study of Swedish women: the Women's Health in the Lund Area (WHILA) Study. Fertil Steril 88:163–171
4. **Ding EL, Song Y, Manson JE, Rifai N, Buring JE, Liu S** 2007 Plasma

sex steroid hormones and risk of developing type 2 diabetes in women: a prospective study. Diabetologia 50:2076–2084

5. **Kalyani RR, Franco M, Dobs AS, Ouyang P, Vaidya D, Bertoni A, Gapstur SM, Golden SH** 2009 The association of endogenous sex hormones, adiposity, and insulin resistance with incident diabetes in postmenopausal women. J Clin Endocrinol Metab 94:4127–4135
6. **Shahar E, Burke GL, Cushman M, Heckbert SR, Ouyang P, Szklo M** 2008 Post menopausal hormones and measures of subclinical atherosclerosis: the multi-ethnic study of atherosclerosis. Prev Med 47: 38–45
7. **Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, Jackson RD, Beresford SA, Howard BV, Johnson KC, Kotchen JM, Ockene J** 2002 Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA 288:321–333
8. **Laughlin GA, Goodell V, Barrett-Connor E** 2010 Extremes of endogenous testosterone are associated with increased risk of incident coronary events in older women. J Clin Endocrinol Metab 95:740–747
9. **Ding EL, Song Y, Manson JE, Hunter DJ, Lee CC, Rifai N, Buring JE, Gaziano JM, Liu S** 2009 Sex hormone-binding globulin and risk of type 2 diabetes in women and men. N Engl J Med 361:1152–1163
10. **Perry JR, Weedon MN, Langenberg C, Jackson AU, Lyssenko V, Sparsø T, Thorleifsson G, Grallert H, Ferrucci L, Maggio M, Paolisso G, Walker M, Palmer CN, Payne F, Young E, Herder C, Narisu N, Morken MA, Bonnycastle LL, Owen KR, Shields B, Knight B, Bennett A, Groves CJ, Ruokonen A,** *et al.* 2010 Genetic evidence that raised sex hormone binding globulin (SHBG) levels reduce the risk of type 2 diabetes. Hum Mol Genet 19:535–544
11. **Maggio M, Lauretani F, Basaria S, Ceda GP, Bandinelli S, Metter EJ, Bos AJ, Ruggiero C, Ceresini G, Paolisso G, Artoni A, Valenti G, Guralnik JM, Ferrucci L** 2008 Sex hormone binding globulin levels across the adult lifespan in women: the role of body mass index and fasting insulin. J Endocrinol Invest 31:597–601
12. **Jaquish CE, Blangero J, Haffner SM, Stern MP, MacCluer JW** 1997 Quantitative genetics of serum sex hormone-binding globulin levels in participants in the San Antonio Family Heart Study. Metabolism 46:988–991
13. **Meikle AW, Bishop DT, Stringham JD, West DW** 1986 Quantitating genetic and nongenetic factors that determine plasma sex steroid variation in normal male twins. Metabolism 35:1090–1095
14. **Ring HZ, Lessov CN, Reed T, Marcus R, Holloway L, Swan GE, Carmelli D** 2005 Heritability of plasma sex hormones and hormone binding globulin in adult male twins. J Clin Endocrinol Metab 90: 3653–3658
15. **Stone J, Folkerd E, Doody D, Schroen C, Treloar SA, Giles GG, Pike MC, English DR, Southey MC, Hopper JL, Dowsett M** 2009 Familial correlations in postmenopausal serum concentrations of sex steroid hormones and other mitogens: a twins and sisters study. J Clin Endocrinol Metab 94:4793–4800
16. **Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino Sr RB, Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D** 2007 The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol 165:1328–1335
17. **Sir-Petermann T, Codner E, Pérez V, Echiburú B, Maliqueo M, Ladrón de Guevara A, Preisler J, Crisosto N, Sánchez F, Cassorla F, Bhasin S** 2009 Metabolic and reproductive features before and during puberty in daughters of women with polycystic ovary syndrome. J Clin Endocrinol Metab 94:1923–1930
18. **Mazer NA** 2009 A novel spreadsheet method for calculating the free serum concentrations of testosterone, dihydrotestosterone, estradiol, estrone and cortisol: with illustrative examples from male and female populations. Steroids 74:512–519
19. **Almasy L, Blangero J** 1998 Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62:1198–1211
20. **Davison SL, Bell R, Donath S, Montalto JG, Davis SR** 2005 Androgen levels in adult females: changes with age, menopause, and oophorectomy. J Clin Endocrinol Metab 90:3847–3853
21. **Agrinier N, Cournot M, Dallongeville J, Arveiler D, Ducimetière P, Ruidavets JB, Ferrières J** 2010 Menopause and modifiable coronary heart disease risk factors: a population based study. Maturitas 65: 237–243
22. **Coviello AD, Vasan RS, Pencina M, D'Agostino R, Zhang A, Ulloor J, Bhasin S** 2007 Free testosterone is positively associated with diabetes, metabolic syndrome, and dyslipidemia in women: the Framingham Offspring Study. J Womens Health 16:1113
23. **Shaw LJ, Bairey Merz CN, Azziz R, Stanczyk FZ, Sopko G, Braunstein GD, Kelsey SF, Kip KE, Cooper-Dehoff RM, Johnson BD, Vaccarino V, Reis SE, Bittner V, Hodgson TK, Rogers W, Pepine CJ** 2008 Postmenopausal women with a history of irregular menses and elevated androgen measurements at high risk for worsening cardiovascular event-free survival: results from the National Institutes of Health—National Heart, Lung, and Blood Institute sponsored Women's Ischemia Syndrome Evaluation. J Clin Endocrinol Metab 93:1276–1284
24. **Hryb DJ, Khan MS, Romas NA, Rosner W** 1990 The control of the interaction of sex hormone-binding globulin with its receptor by steroid hormones. J Biol Chem 265:6048–6054
25. **Hammes A, Andreassen TK, Spoelgen R, Raila J, Hubner N, Schulz H, Metzger J, Schweigert FJ, Luppa PB, Nykjaer A, Willnow TE** 2005 Role of endocytosis in cellular uptake of sex steroids. Cell 122:751–762

ELSEVIER

Journal of Substance Abuse Treatment 41 (2011) 179–185

Journal of Substance Abuse Treatment

Regular article

Oral health of substance-dependent individuals: Impact of specific substances

Meredith M. D'Amore, (M.P.H.)[a,*], Debbie M. Cheng, (Sc.D.)[b,c,d], Nancy R. Kressin, (Ph.D.)[a,e,f], Judith Jones, (D.D.S., M.P.H., DSc.D.)[g], Jeffrey H. Samet, (M.D., M.A., M.P.H.)[b,c,h], Michael Winter, (M.P.H.)[i], Theresa W. Kim, (M.D.)[b,c], Richard Saitz, (M.D., M.P.H.)[b,c,j]

[a]*Health/care Disparities Research Program, Section of General Internal Medicine, Boston University School of Medicine, Boston Medical Center, Boston, MA, 02118, USA*
[b]*Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, 02118, USA*
[c]*Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston University School of Medicine and Boston Medical Center, Boston, MA, 02118, USA*
[d]*Department of Biostatistics, Boston University School of Public Health, Boston, MA, 02118, USA*
[e]*VA Boston Healthcare System, Boston, MA, 02130, USA*
[f]*Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, Bedford, MA, 01730, USA*
[g]*Boston University Henry M. Goldman School of Dental Medicine, Boston, MA, 02118, USA*
[h]*Department of Community Health Sciences, Boston University School of Public Health, Boston, MA, 02118, USA*
[i]*Data Coordinating Center, Boston University School of Public Health, Boston, MA, 02118, USA*
[j]*Department of Epidemiology, Boston University School of Public Health, Boston, MA, 02118, USA*

Received 21 October 2010; received in revised form 3 February 2011; accepted 7 February 2011

Abstract

Little is known about how different types of substances affect oral health. Our objective was to examine the respective effects of alcohol, stimulants, opioids, and marijuana on oral health in substance-dependent persons. Using self-reported data from 563 substance-dependent individuals, we found that most reported unsatisfactory oral health, with their most recent dental visit more than 1 year ago. In multivariable logistic regressions, none of the substance types were significantly associated with oral health status. However, opioid use was significantly related to a worse overall oral health rating compared to 1 year ago. These findings highlight the poor oral health of individuals with substance dependence and the need to address declining oral health among opioid users. General health and specialty addiction care providers should be aware of oral health problems among these patients. In addition, engagement into addiction and medical care may be facilitated by addressing oral health concerns.

Keywords: Oral health; Substance dependence; Dental care

1. Introduction

Poor oral health is a common problem among individuals with substance dependence, yet this topic has been largely neglected in the addiction literature (Reece, 2009). Individuals with heavy substance use are at increased risk for poor oral health for a variety of reasons, including limited access to dental care (Johnson, Hearn, & Barker, 2008; Khocht, Schleifer, Janal, & Keller, 2009; Sheridan, Aggleton, & Carson, 2001; ter Horst, Molendijk, Brouwer, & Verhey,

This project was supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) Grant R01AA010870, National Institute on Drug Abuse (NIDA) Grant R01DA010019, and National Center for Research Resources (NCRR) Grant UL1RR025771. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIAAA, NIDA, NCRR, or the National Institutes of Health.

Dr. Jones is supported by K24DE018211.

* Corresponding author. 801 Massachusetts Avenue, #2098, Boston, MA 02113. Tel.: +1 617 414 8441; fax: +1 617 638 8026.

E-mail address: meredith.damore@bmc.org (M.M. D'Amore).

0740-5472/11/$ – see front matter
doi:10.1016/j.jsat.2011.02.005

1996), poor dietary (Laslett, Dietze, & Dwyer, 2008; Morio, Marshall, Qian, & Morgan, 2008; Titsas and Ferguson, 2002) and oral hygiene habits (Barbadoro, Lucrezi, Prospero, & Annino, 2008; Friedlander, Marder, Pisegna, & Yagiela, 2003; Morio et al., 2008), negative attitudes about oral health and health care (Robinson, Acquah, & Gibson, 2005), and direct physical effects of the substance on oral health. There are several mechanisms by which drugs can directly affect oral health, including increased xerostomia (dry mouth) due to hyposalivation (lack of salivary flow), poor diet and self-care leading to higher rates of dental caries, enamel erosion, and periodontal disease (Friedlander et al., 2003; Hamamoto and Rhodus, 2009; Morio et al., 2008; Versteeg, Slot, van der Velden, & van der Weijden, 2008).

Prior research suggests that individuals who abuse methamphetamines (Curtis, 2006; Donaldson and Goodchild, 2006; Hamamoto and Rhodus, 2009; Morio et al., 2008; Shetty et al., 2010), alcohol (Araujo, Dermen, Connors, & Ciancio, 2004; Hornecker, Muuss, Ehrenreich, & Mausberg, 2003; Khocht et al., 2009; Manarte, Manso, Souza, Frias-Bulhosa, & Gago, 2009), opioids (Sheedy, 1996; Steinmiller and Greenwald, 2007), marijuana (Versteeg et al., 2008), and cocaine (Brand, Gonggrijp, & Blanksma, 2008) are at increased risk of poor oral health outcomes, including enamel erosion and caries. It is not clear, however, whether these consequences are substance specific or due to substance dependence in general. Suboptimal oral health and periodontal disease are associated with health consequences localized to dental issues, such as tooth loss (Martin, Page, Loeb, & Levi, 2010), and more pervasive physical health problems (Slots, 2003), including cerebrovascular disease (Wu et al., 2000), low birth weight (Cruz et al., 2009; Moliterno, Monteiro, Figueredo, & Fischer, 2005), pulmonary infection (Mojon, 2002), diabetes (Demmer, Jacobs, & Desvarieux, 2008), and potentially cardiovascular disease (Humphrey, Fu, Buckley, Freeman, & Helfand, 2008). Given the deleterious impact of poor oral health on both local and systemic health outcomes, it is critical to identify populations at increased risk in an effort to develop tailored interventions to improve overall health in this population.

Despite its prevalence and consequences, there have been few studies that compare and contrast the oral health of patients who use different substances. Most studies that have examined oral health among substance users are cross-sectional or case studies and have assessed only one type of substance (Araujo et al., 2004; Hornecker et al., 2003; Khocht et al., 2009; Manarte et al., 2009; Morio et al., 2008; Versteeg et al., 2008; Shetty et al., 2010) and include small samples (Araujo et al., 2004; Johnson et al., 2008; Khocht et al., 2009; Morio et al., 2008; Robinson et al., 2005; Sheedy, 1996). Many studies and reviews that have addressed the issue of substance use and oral health were conducted internationally, with little work focused in the United States, where the stigma of substance use and perceptions of and access to treatment for oral health may be different (Barbadoro et al., 2008; Blanksma and Brand, 2004; Cho, Hirsch, & Johnstone, 2005; Johnson et al., 2008; Laslett et al., 2008; Molendijk, Ter Horst, Kasbergen, Truin, & Mulder, 1996; Pilinova, Krutina, Salandova & Pilin, 2003; Reece, 2007; Robinson et al., 2005; Sheridan et al., 2001; ter Horst et al., 1996). Understanding how specific types of substances affect oral health can potentially help target interventions to certain groups at risk. Thus, the purpose of this study is to examine whether substance use, including alcohol, stimulants, opioids, and marijuana, is associated with oral health status among people with substance dependence.

2. Materials and methods

2.1. Design

We analyzed data on self-rated oral health and substance use collected prospectively from participants enrolled in a randomized trial testing the effectiveness of chronic care management in the primary care setting in Boston from September 2006 to September 2008. This study was approved by the Boston University Medical Campus Institutional Review Board. All subjects provided informed consent, and procedures were followed in accordance with the Helsinki Declaration of 1975. A certificate of confidentiality was obtained from the National Institute on Alcohol Abuse and Alcoholism to further protect participants' privacy. We hypothesized that different types of substances would differentially affect oral health.

2.2. Sample

The sample included 563 men and women who reported using a variety of substances and had enrolled in the Addiction Health Evaluation and Disease Management study. All subjects had current alcohol and/or drug dependence by *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* criteria (American Psychiatric Association, 2000) assessed using the Composite International Diagnostic Interview Short form (Gigantesco and Morosini, 2008) and were willing to establish or continue primary medical care at the study location. Participants were included if they reported past 30-day drug use or heavy alcohol use (defined as ≥4 standard drinks for women, ≥5 for men at least twice; or >14 drinks per week for women, >21 drinks per week for men, in an average week in the past month). Approximately 74% of the subjects were recruited from a detoxification center, 9% from ambulatory care/outpatient settings, 1% from the hospital emergency department or inpatient setting, and 16% from other sources. Subjects were at least 18 years of age, spoke English or Spanish, and were without indication of cognitive impairment at screening (assessed using the Mini Mental State Examination score greater than 20; Klein et al., 1985). For this analysis, data were taken from an interview conducted at study entry prior to randomization; thus, any intervention effects would not impact survey responses.

2.3. Measures

2.3.1. Independent variables: Types of substances used

The four main independent variables of interest represented use of a particular type of substance: heavy alcohol (yes/no), stimulants (yes/no), opioids (yes/no), and marijuana (yes/no). All subjects in this study met criteria for substance dependence; however, for this analysis, we are analyzing different types of substance use. Thus, the term *substance use* is used to refer to these variables. Heavy alcohol use was assessed using the timeline follow-back measure and defined as drinking five or more drinks at least 1 day in the past 30 days (if male) or four or more drinks at least 1 day in the past 30 days (if female; National Institute on Alcohol Abuse and Alcoholism, 2005). Other drug use was assessed using the Addiction Severity Index (McLellan, Luborsky, Woody, & O'Brien, 1980). Stimulant use was defined as using cocaine or amphetamines at least 1 day in the past 30 days. Individuals who used heroin, methadone, or other opioid analgesics, either without a doctor's prescription, in larger amounts than prescribed, or for a longer period than prescribed, at least 1 day in the past 30 days were considered opioid users. Marijuana use was defined as using marijuana or cannabis at least 1 day in the past 30 days.

2.3.2. Dependent variables: Oral health indicators

Our primary outcome was self-reported oral health status ("How would you describe the health of your teeth and gums?"; Jones et al., 2004). Subjects rated this item on a 5-point Likert scale, which was then dichotomized into *satisfactory* oral health, defined as "excellent," "very good," or "good," versus *unsatisfactory*, defined as "fair" or "poor" (Cunha-Cruz, Hujoel, & Kressin, 2007; Jones, Spiro, Miller, Garcia, & Kressin, 2002). Four secondary self-reported outcomes related to oral health were also evaluated. Health of teeth and gums compared to 1 year ago was dichotomized as *worse*, defined as "somewhat" or "much" worse, versus *the same*, "somewhat" or "much" better. Tooth or gum pain in the past 3 months was dichotomized as *pain* ("some," "quite a bit," or "a great deal") versus *no pain* ("little" or "none"). The time since last dentist visit was dichotomized as *recent* (<1 year ago) versus *distant* (≥1 year ago, never been to the dentist, or "don't know"). The number of permanent teeth removed because of tooth decay, gum disease, or infection was dichotomized as 0–5 versus 6 or more (Kapp, Boren, Yun, & LeMaster, 2007) The rationale for this dichotomy was that individuals could have had all four wisdom teeth or bicuspids extracted for orthodontic treatment and still have excellent oral health.

2.4. Covariates

The analyses controlled for sociodemographic and other variables that could affect oral health, including age, gender, education, being a current smoker, race/ethnicity, health insurance, and income. Health status was assessed using one item of a general rating of health on a 5-point Likert scale ranging from "excellent" to "poor", analyzed as a continuous variable (Ware, Kosinski, & Keller, 1996). In addition, oral health can also be influenced by lifestyle, including dietary and hygiene habits. Individuals in prison (Walsh, Tickle, Milsom, Buchanan, & Zoitopoulos, 2008) or who have experienced homelessness (De Palma and Nordenram, 2005; Gibson et al., 2003) may be at an increased risk for worse oral health due to poorer lifestyle habits and access to care. Thus, we included covariates of ever spending time in prison and recent homelessness, the latter defined as spending at least one night in a shelter or on the street in the last 3 months. Polysubstance use was also included and defined as using two or more of the above substances (heavy alcohol, stimulants, opioids, or marijuana) in the last 30 days.

2.5. Statistical analyses

Descriptive statistics were used to assess the bivariate relationship between subject characteristics and the primary outcome of unsatisfactory oral health. Two sample *t* tests and chi-square tests were used as appropriate to assess the bivariate associations. We evaluated the association between the types of substances used and each oral health outcome using separate logistic regression models. The multivariable logistic regression models were fit to evaluate the associations between types of substances used and worse oral health outcomes after adjustment for sociodemographic characteristics, health status, and lifestyle variables. Adjusted odds ratios (AORs) and 95% confidence intervals (CIs) are reported. To minimize the potential for collinearity, we assessed correlation between pairs of independent variables and covariates to identify pairs of variables that were correlated (i.e., $r > .50$). Polysubstance use was moderately correlated with stimulant use ($r = .66$). Polysubstance use was expected to be an important factor and potential confounder; therefore, we fit adjusted models with and without this covariate. Adjustment for polysubstance use attenuated the odds ratios for the main independent variables with oral health status. Thus, we present results controlling for polysubstance use as the final models. All analyses were conducted using two-sided tests and a significance level of .05. Analyses were performed using SAS software (version 9.1; SAS Institute, Cary, NC).

3. Results

The majority of our sample was male (73%), currently smoking (88%), less than 50 years of age (88%; mean age = 38), with an annual income less than $50,000 (74%), had at least a high school education (76%), and had some form of health insurance (79%; Table 1). Almost half were White (47%), 32% Black, 13% Hispanic, and 8% other racial/ethnic background. Most were incarcerated at least once during

Table 1
Sample characteristics by self-rated oral health status

Characteristic	All subjects (n = 563)	Subjects with fair or poor self-rated oral health (n = 335)	Subjects with good, very good, or excellent self-rated oral health (n = 228)	p
Age at baseline, n (%)				
18–34	**212 (38)**	**111 (33)**	**101 (44)**	**.01**
35–49	**283 (50)**	**174 (52)**	**109 (48)**	
50+	**68 (12)**	**50 (15)**	**18 (8)**	
Age, M (SD)	**38 (10)**	**39 (10)**	**37 (10)**	**<.01**
Gender, n (%) male	409 (73)	241 (72)	168 (74)	.65
Education, n (%)				
<High school	133 (24)	89 (27)	44 (19)	.12
High school graduate	277 (49)	156 (47)	121 (53)	
>High school	153 (27)	90 (27)	63 (28)	
Current smoker	**493 (88)**	**302 (90)**	**191 (84)**	**.02**
Race, n (%)				
White	264 (47)	153 (46)	111 (49)	.57
Black	179 (32)	114 (34)	65 (29)	
Hispanic	75 (13)	43 (13)	32 (14)	
Other	45 (8)	25 (8)	20 (9)	
Any health insurance	446 (79)	269 (80)	177 (78)	.50
Income, n (%)				
<$20,000	236 (42)	153 (46)	83 (36)	.08
$20,000–$49,999	178 (32)	99 (30)	79 (35)	
$50,000+	147 (26)	81 (24)	66 (29)	
Overall health status, M (SD)[a]	**3.0 (1.0)**	**3.2 (1.0)**	**2.7 (1.0)**	**<.01**
Ever incarcerated, n (%) yes	438 (78)	254 (76)	184 (81)	.14
Homeless, n (%) yes	332 (59)	199 (59)	133 (58)	.80
Heavy alcohol use, n (%) yes	440 (78)	270 (81)	170 (75)	.09
Stimulant use, n (%) yes	382 (68)	230 (69)	152 (67)	.62
Opioid use, n (%) yes	378 (67)	223 (67)	155 (68)	.73
Marijuana use, n (%) yes	275 (49)	159 (48)	116 (51)	.43
Polysubstance use, n (%) yes	458 (81)	279 (83)	179 (79)	.15

Note. Bold indicates $p < .05$.

[a] Range = 1–5, where 1 is *excellent* and 5 is *poor*.

their lives (78%), and many noted homelessness in the past 3 months (59%). Most reported heavy alcohol (78%), stimulant use (68%), opioid use (67%), and polysubstance use (81%). About half used marijuana in the last month (49%).

Overall, the majority of the sample reported unsatisfactory oral health (60%), with the most recent dental visit being more than 1 year ago or not able to recall (52%; Table 2). However, most reported the same or better oral health compared with 1 year ago (64%), little or no tooth/gum pain (63%), and having less than six teeth removed (71%).

3.1. Primary outcome: Oral health status

In unadjusted analyses, there were no significant associations between heavy alcohol use, stimulant use, opioid use, or marijuana use and unsatisfactory self-rated oral health. The findings were similar in adjusted analyses, where those with heavy alcohol use (AOR = 1.31, 95% CI = 0.79–2.15) and opioid use (AOR = 1.09, 95% CI = 0.66–1.79) had nonsignificant higher odds of unsatisfactory self-reported oral health; those with stimulant use (AOR = 0.85, 95% CI = 0.49–1.47) and marijuana use (AOR = 0.69, 95% CI = 0.45–1.06) had nonsignificant lower odds of unsatisfactory oral health (Table 3). In multivariable models, individuals who were older (AOR = 2.86, 95% CI = 1.41–5.80), currently smoking (AOR = 1.99, 95% CI = 1.13–3.50), and had a worse overall health status rating (AOR =

Table 2
Self-reported oral health outcomes (n = 563)

Item	n (%)
How would you describe the health of your teeth and gums?	
Fair or poor	335 (60)
Good, very good, or excellent	228 (40)
Compared with 1 year ago, how would you rate the health of your teeth and gums today?	
Somewhat or much worse	201 (36)
The same, somewhat, or much better	362 (64)
During the past 3 months, how much pain or distress have your teeth or gums caused you?	
Some, quite a bit, or a great deal	209 (37)
A little bit or none	354 (63)
About how long has it been since you last saw a dentist?	
More than 1 year ago, never, don't know	292 (52)
1 year ago or less	271 (48)
How many of your permanent teeth have been removed because of tooth decay or gum disease?	
6 or more	161 (29)
5 or fewer	387 (71)

Table 3
Adjusted models evaluating the associations between substance use types and oral health outcomes[a]

Substance use types	Oral health outcomes									
	Unsatisfactory oral health status		Worse oral health compared with 1 year ago		Tooth/Gum pain		>1 year since last dentist visit		≥6 teeth removed	
	AOR (95% CI)	*p*	AOR (95% CI)	*p*	AOR (95% CI)	*p*	AOR (95% CI)	*p*	AOR (95% CI)	*p*
Heavy alcohol use										
Yes vs. No	1.31 (0.79–2.15)	.29	1.13 (0.68–1.87)	.63	1.02 (0.62–1.67)	.95	1.16 (0.71–1.89)	.55	0.58 (0.32–1.05)	.07
Stimulant use										
Yes vs. No	0.85 (0.49–1.47)	.56	0.99 (0.58–1.70)	.98	1.15 (0.67–1.97)	.60	1.19 (0.71–2.01)	.51	1.21 (0.62–2.39)	.57
Opioid use										
Yes vs. No	1.09 (0.66–1.79)	.74	**1.72 (1.04–2.82)**	**.03**	1.54 (0.95–2.52)	.08	0.92 (0.58–1.48)	.74	1.26 (0.73–2.15)	.41
Marijuana use										
Yes vs. No	0.69 (0.45–1.06)	.09	0.76 (0.50–1.15)	.19	0.81 (0.54–1.22)	.32	0.77 (0.51–1.16)	.22	0.76 (0.46–1.26)	.29

Note. Bold indicates $p < .05$.

[a] Separate models were fit for each outcome, each adjusting for age, education, gender, race/ethnicity, smoking, health insurance, income, health status, lifetime incarceration, homelessness, and polysubstance use.

1.66, 95% CI = 1.37–2.01) had significantly higher odds of unsatisfactory oral health status (not shown).

3.2. Secondary outcomes

In adjusted analyses, opioid use was the only substance significantly related to a worse oral health rating compared with 1 year ago (AOR = 1.72, CI = 1.04–2.82; Table 3). None of the substance types were significantly associated with tooth/gum pain, time since last dental visit, or number of teeth removed. Overall, in adjusted analyses across secondary outcomes, marijuana use was associated with lower odds of worse oral health outcomes, although none of these associations were statistically significant.

4. Discussion

Overall, most of these individuals with substance dependence reported unsatisfactory oral health status, consistent with previous literature (Araujo et al., 2004). Contrary to our hypothesis, we did not detect an association between type of substance and self-reported oral health status. Of note, however, is that the reference group for each comparison is people with substance dependence who do not use the particular substance of interest (e.g., those with heavy alcohol use vs. those without heavy alcohol use but who use other substances), and thus, our findings should not be interpreted or generalized to represent the impact of a particular substance compared with no other substance use. Our results did reveal that opioid use is associated with worse self-rated oral health compared with 1 year ago. This is consistent with research that has found an association between opiate use and poor oral health (Nathwani and Gallagher, 2008; Reece, 2007; Sheedy, 1996; Titsas and Ferguson, 2002). The association could be due to direct effects of opioids, or it could be that people with worsening oral health use opioids for relief (although this seems less likely given the absence of an association between opioid use and dental pain). Of note, however, is that opioid use is reported for the previous 30 days, and worse oral health is compared with 1 year ago. Subjects may be disappointed in their continued substance use, which may have affected their reports of worsening oral health status.

The association between use of methamphetamines or cocaine and poor oral health is well established in the literature (Brand et al., 2008; Curtis, 2006; Donaldson and Goodchild, 2006; Hamamoto and Rhodus, 2009; Shetty et al., 2010). Similarly, research indicates there is an association between alcohol dependence or marijuana use and poor oral health (Araujo et al., 2004; Hornecker et al., 2003; Khocht et al., 2009; Manarte et al., 2009; Versteeg et al., 2008). These findings from other studies are in contrast to ours. One possibility regarding stimulants is that although the prevalence of cocaine use was substantial in our sample, the prevalence of methamphetamine use was not. If more effects on oral health would be seen from the latter, then that could explain the absence of effect. However, one would have expected effects from cocaine itself. Another possible explanation for different findings is that most of the prior studies relied on oral examination to evaluate oral health outcomes. Our findings, that no individual substance type was related to overall oral health status, may be a function of individual subjectivity in the assessment of their oral health. Although methamphetamine use is associated with poor oral health compared with no methamphetamine use, it may not be associated with poor oral health more than other substances in people with dependence. Similarly, given the large percentage of polysubstance use in our sample, we may not have been able to disentangle the effects of specific substances on oral health.

Twenty-nine percent of our sample had six or more teeth removed. This is substantially more than the 8.5% of adults in the general population who have had six or more teeth removed, according to the Behavioral Risk Factor Surveillance Survey data from 2004 (Kapp et al., 2007). This finding

confirms other research documenting worse oral health status among substance-dependent populations (Reece, 2009).

Although not the focus of this analysis, we noted that some covariates behaved as we expected, supporting their inclusion in the adjusted models. The association between smoking and worse oral health is not surprising, given the previous literature confirming their correlation (Friedlander et al., 2003; Morio et al., 2008). The persistent association between lower self-rated health status and oral health across several models suggests that oral health may influence physical health or vice versa. A longitudinal study examining health status over time could better determine that causal pathway. Future research should also compare self-reported oral health of this population to the general population to understand the extent to which oral health is worse among individuals with substance dependence.

These findings should be interpreted within the limitations of our analysis. All of our oral health outcomes are self-reported and thus subject to recall and other potential biases; we did not conduct a clinical dental health examination. We believe there is value in understanding patients' perceptions of their oral health, given that subjective assessment can affect health behaviors (Baker, 2009; Kneckt, Syrjala, & Knuuttila, 1999). For number of teeth removed, which can be considered a less subjective measure, there is evidence that self-report is highly correlated with the number of teeth found missing on clinical examination and thus can be considered reasonably reliable (Pitiphat, Garcia, Douglass, & Joshipura, 2002). We included a measure of recent homelessness in our analysis; however, an estimate of long-term homelessness may have been more appropriate considering its chronic impact on lifestyle, dietary habits, and oral health care. Our definition of substance use was defined within the last 30 days, which may not entirely capture the effects on oral health, given that long-term exposure of a particular substance would have more of an impact on these outcomes. Given our reference group in adjusted analyses, we are not able to assess the independent effects of a particular substance compared with people who did not use any substances. Instead, our results indicate the effects of a substance on oral health, above and beyond the effects of other substances. Finally, we do not have data to compare to non-substance-using populations, and thus, it is difficult to tease out the effects of substance use and other sociodemographic characteristics. We attempted to do so in our adjusted model, controlling for age, education, gender, race/ethnicity, smoking, health insurance, income, health status, lifetime incarceration, homelessness, and polysubstance use.

Despite these limitations, this study is among the first to assess the differential effects of varying types of substance use on self-rated oral health outcomes. Our findings suggest that addiction treatment providers, as well as medical and dental clinicians, should consider dental and addiction problems as associated comorbidities, requiring the development of treatment plans that address both substance use and potential oral health problems. These results suggest that type of substance had little effect on oral health outcomes. However, the overall poor rating of oral health in our sample indicates that the health of teeth and gums is a significant issue among individuals who use alcohol and/or drugs, in general. Thus, interventions for poor oral health could be tailored toward this population.

Despite the association between substance use and oral health and national recommendations for improving oral health among such vulnerable populations (Department of Health and Human Services, 2000), few interventions have been targeted to individuals with substance dependence. One intervention aimed at improving knowledge, attitudes, and behavior of individuals with alcohol dependence found that the intervention group who attended a lecture had significant improvement in oral health behaviors after 1 year, including frequency of tooth brushing (Barbadoro et al., 2008). Although conducted in Italy, this study has implications for rehabilitation programs in the United States, where such educational workshops can be incorporated into treatment. Clinicians should pay particular attention to oral health among opioid users. Given our findings, and the correlation between oral and general health status, oral health warrants increased attention and public health efforts among substance users. In addition, general health and specialty addiction care providers should be aware of oral health problems among patients with substance dependence. Finally, engagement into addiction and medical care, often a challenge in this population, may be facilitated by addressing oral health concerns.

Acknowledgments

No authors have identified any financial conflicts of interest for this study.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders.* Washington, DC: American Psychiatric Association.

Araujo, M. W., Dermen, K., Connors, G., & Ciancio, S. (2004). Oral and dental health among inpatients in treatment for alcohol use disorders: A pilot study. *Journal of the International Academy of Periodontology*, *6*, 125−130.

Baker, S. R. (2009). Applying Andersen's behavioural model to oral health: What are the contextual factors shaping perceived oral health outcomes? *Community Dentistry and Oral Epidemiology*, *37*, 485−494.

Barbadoro, P., Lucrezi, D., Prospero, E., & Annino, I. (2008). Improvement of knowledge, attitude, and behavior about oral health in a population of alcohol addicted persons. *Alcohol and Alcoholism*, *43*, 347−350.

Blanksma, C. J., & Brand, H. S. (2004). Effects of cocaine use on oral health and implications for dental treatment. *Nederlands Tijdschrift Voor Tandheelkunde*, *111*, 486−489.

Brand, H. S., Gonggrijp, S., & Blanksma, C. J. (2008). Cocaine and oral health. *British Dental Journal*, *204*, 365−369.

Cho, C. M., Hirsch, R., & Johnstone, S. (2005). General and oral health implications of cannabis use. *Australian Dental Journal*, *50*, 70−74.

Cruz, S. S., Costa Mda, C., Gomes-Filho, I. S., et al. (2009). Contribution of periodontal disease in pregnant women as a risk factor for low birth weight. *Community Dentistry and Oral Epidemiology*, *37*, 527−533.

Cunha-Cruz, J., Hujoel, P. P., & Kressin, N. R. (2007). Oral health-related quality of life of periodontal patients. *Journal of Periodontal Research*, *42*, 169−176.

Curtis, E. K. (2006). Meth mouth: A review of methamphetamine abuse and its oral manifestations. *General Dentistry*, *54*, 125−129 [quiz 130].

Demmer, R. T., Jacobs, D. R., Jr., & Desvarieux, M. (2008). Periodontal disease and incident type 2 diabetes: Results from the First National Health and Nutrition Examination Survey and its epidemiologic follow-up study. *Diabetes Care*, *31*, 1373−1379.

De Palma, P., & Nordenram, G. (2005). The perceptions of homeless people in Stockholm concerning oral health and consequences of dental treatment: A qualitative study. *Special Care in Dentistry*, *25*, 289−295.

Department of Health and Human Services. (2000). *Oral health in America: A report of the surgeon general.* Rockville, MD, U.S: Department of Health and Human Service, National Institute of Dental and Craniofacial Research, National Institutes of Health.

Donaldson, M., & Goodchild, J. H. (2006). Oral health of the methamphetamine abuser. *American Journal of Health-System Pharmacy*, *63*, 2078−2082.

Friedlander, A. H., Marder, S. R., Pisegna, J. R., & Yagiela, J. A. (2003). Alcohol abuse and dependence: Psychopathology, medical management and dental implications. *Journal of the American Dental Association*, *134*, 731−740.

Gibson, G., Rosenheck, R., Tullner, J. B., et al. (2003). A national survey of the oral health status of homeless veterans. *Journal of Public Health Dentistry*, *63*, 30−37.

Gigantesco, A., & Morosini, P. (2008). Development, reliability and factor analysis of a self-administered questionnaire which originates from the World Health Organization's Composite International Diagnostic Interview—Short Form (CIDI-SF) for assessing mental disorders. *Clinical Practice in Epidemiology and Mental Health*, *4*, 8.

Hamamoto, D. T., & Rhodus, N. L. (2009). Methamphetamine abuse and dentistry. *Oral Diseases*, *15*, 27−37.

Hornecker, E., Muuss, T., Ehrenreich, H., & Mausberg, R. F. (2003). A pilot study on the oral conditions of severely alcohol addicted persons. *Journal of Contemporary Dental Practice*, *4*, 51−59.

Humphrey, L. L., Fu, R., Buckley, D. I., Freeman, M., & Helfand, M. (2008). Periodontal disease and coronary heart disease incidence: A systematic review and meta-analysis. *Journal of General Internal Medicine*, *23*, 2079−2086.

Johnson, D., Hearn, A., & Barker, D. (2008). A pilot survey of dental health in a group of drug and alcohol abusers. *European Journal of Prosthodontics and Restorative Dentistry*, *16*, 181−184.

Jones, J. A., Spiro, A., III, Miller, D. R., Garcia, R. I., & Kressin, N. R. (2002). Need for dental care in older veterans: Assessment of patient-based measures. *Journal of the American Geriatrics Society*, *50*, 163−168.

Jones, J., Kressin, N., Miller, D., et al. (2004). Comparison of patient-based oral health outcome measures. *Quality of Life Research*, *13*, 975−985.

Kapp, J. M., Boren, S. A., Yun, S., & LeMaster, J. (2007). Diabetes and tooth loss in a national sample of dentate adults reporting annual dental visits. *Preventing Chronic Disease*, *4*, A59.

Khocht, A., Schleifer, S. J., Janal, M. N., & Keller, S. (2009). Dental care and oral disease in alcohol-dependent persons. *Journal of Substance Abuse Treatment*, *37*, 214−218.

Klein, L. E., Roca, R. P., McArthur, J., et al. (1985). Diagnosing dementia. Univariate and multivariate analyses of the mental status examination. *Journal of the American Geriatrics Society*, *33*, 483−488.

Kneckt, M. C., Syrjala, A. M., & Knuuttila, M. L. (1999). Locus of control beliefs predicting oral and diabetes health behavior and health status. *Acta Odontologica Scandinavica*, *57*, 127−131.

Laslett, A. M., Dietze, P., & Dwyer, R. (2008). The oral health of street-recruited injecting drug users: Prevalence and correlates of problems. *Addiction*, *103*, 1821−1825.

Manarte, P., Manso, M. C., Souza, D., Frias-Bulhosa, J., & Gago, S. (2009). Dental erosion in alcoholic patients under addiction rehabilitation therapy. *Medicina Oral, Patologia Oral y Cirugia Bucal*, *14*, e376−e383.

Martin, J. A., Page, R. C., Loeb, C. F., & Levi, P. A., Jr (2010). Tooth loss in 776 treated periodontal patients. *Journal of Periodontology*, *81*, 244−250.

McLellan, A. T., Luborsky, L., Woody, G. E., & O'Brien, C. P. (1980). An improved diagnostic evaluation instrument for substance abuse patients. The Addiction Severity Index. *Journal of Nervous and Mental Disease*, *168*, 26−33.

Mojon, P. (2002). Oral health and respiratory infection. *Journal (Canadian Dental Association)*, *68*, 340−345.

Molendijk, B., Ter Horst, G., Kasbergen, M., Truin, G. J., & Mulder, J. (1996). Dental health in Dutch drug addicts. *Community Dentistry and Oral Epidemiology*, *24*, 117−119.

Moliterno, L. F., Monteiro, B., Figueredo, C. M., & Fischer, R. G. (2005). Association between periodontitis and low birth weight: A case-control study. *Journal of Clinical Periodontology*, *32*, 886−890.

Morio, K. A., Marshall, T. A., Qian, F., & Morgan, T. A. (2008). Comparing diet, oral hygiene and caries status of adult methamphetamine users and nonusers: A pilot study. *Journal of the American Dental Association*, *139*, 171−176.

Nathwani, N. S., & Gallagher, J. E. (2008) Methadone: Dental risks and preventive action, Dent Update, 35, 542-544, 547-548.

National Institute on Alcohol Abuse and Alcoholism. (2005). *A pocket guide for alcohol screening and brief intervention.* Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.

Pilinova, A., Krutina, M., Salandova, M., & Pilin, A. (2003). Oral health status of drug addicts in the Czech Republic. *Journal of Forensic Odonto-Stomatology*, *21*, 36−39.

Pitiphat, W., Garcia, R. I., Douglass, C. W., & Joshipura, K. J. (2002). Validation of self-reported oral health measures. *Journal of Public Health Dentistry*, *62*, 122−128.

Reece, A. S. (2007). Dentition of addiction in Queensland: Poor dental status and major contributing drugs. *Australian Dental Journal*, *52*, 144−149.

Reece, S. (2009). Dental health in addiction. *Australian Dental Journal*, *54*, 185−186.

Robinson, P. G., Acquah, S., & Gibson, B. (2005). Drug users: Oral health-related attitudes and behaviours. *British Dental Journal*, *198*, 219−224 [discussion 214].

Sheedy, J. J. (1996). Methadone and caries. Case reports. *Australian Dental Journal*, *41*, 367−369.

Sheridan, J., Aggleton, M., & Carson, T. (2001). Dental health and access to dental treatment: A comparison of drug users and non-drug users attending community pharmacies. *British Dental Journal*, *191*, 453−457.

Shetty, V., Mooney, L. J., Zigler, C. M., et al. (2010). The relationship between methamphetamine use and increased dental disease. *Journal of the American Dental Association*, *141*, 307−318.

Slots, J. (2003). Update on general health risk of periodontal disease. *International Dental Journal*, *53*(Suppl 3), 200−207.

Steinmiller, C. L., & Greenwald, M. (2007). Factors associated with nonmedical use of prescription opioids among heroin-abusing research volunteers. *Experimental and Clinical Psychopharmacology*, *15*, 492−500.

ter Horst, G., Molendijk, B., Brouwer, E., & Verhey, H. G. (1996). Differences in dental treatment plan and planning for drug-addicted and non-drug-addicted patients. *Community Dentistry and Oral Epidemiology*, *24*, 120−123.

Titsas, A., & Ferguson, M. M. (2002). Impact of opioid use on dentistry. *Australian Dental Journal*, *47*, 94−98.

Versteeg, P. A., Slot, D. E., van der Velden, U., & van der Weijden, G. A. (2008). Effect of cannabis usage on the oral environment: A review. *International Journal of Dental Hygiene*, *6*, 315−320.

Walsh, T., Tickle, M., Milsom, K., Buchanan, K., & Zoitopoulos, L. (2008). An investigation of the nature of research into dental health in prisons: A systematic review. *British Dental Journal*, *204*, 683−689 [discussion 667].

Ware, J. Jr., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Medical care*, *34*, 220−233.

Wu, T., Trevisan, M., Genco, R. J., et al. (2000). Periodontal disease and risk of cerebrovascular disease: The First National Health and Nutrition Examination Survey and its follow-up study. *Archives of Internal Medicine*, *160*, 2749−2755.

REVIEW

THE AMERICAN JOURNAL of MEDICINE®

Stroke Prevention in Atrial Fibrillation: Current Status and Near-Future Directions

Raffaele De Caterina, MD, PhD,[a] Elaine M. Hylek, MD, MPH[b]

[a]Institute of Cardiology and Center of Excellence on Aging, "G. d'Annunzio" University, Chieti, Italy; [b]Department of Medicine, Boston University School of Medicine, Boston, Mass.

ABSTRACT

Prevention of atrial fibrillation-related stroke is an important part of atrial fibrillation management. However, stroke risk is not homogeneous and varies with associated morbidities and risk factors. Risk stratification schemes have been developed that categorize patients' stroke risk into classes based on a combination of risk factors. According to the calculated level of risk, guidelines recommend patients with atrial fibrillation receive antithrombotic therapy either as a vitamin K antagonist or aspirin. Despite recommendations, however, many patients with atrial fibrillation do not receive adequate thromboprophylaxis. We will discuss some of the underlying reasons, in part related to the drawbacks associated with vitamin K antagonists. These highlight the need for new anticoagulants in atrial fibrillation. The novel oral anticoagulants in development may overcome some of the limitations of vitamin K antagonists and address their underuse and safety concerns.

 • *The American Journal of Medicine (2011) 124, 793-799*

KEYWORDS: Anticoagulants; Antithrombotic therapy; Atrial fibrillation; Stroke; Thromboembolism

Atrial fibrillation, the most common sustained cardiac arrhythmia,[1] increases the risk of stroke and thromboembolism by 5-fold.[2] Strokes in atrial fibrillation are generally more severe and associated with greater mortality and disability than strokes from other causes.[3] Therefore, prevention of stroke and thromboembolism is an important part of atrial fibrillation management. However, stroke risk varies widely. Stroke risk-stratification schemes categorize patients' stroke risk into classes based on the presence of risk factors.[4] On this basis, guidelines until now have recommended that patients with atrial fibrillation receive some form of antithrombotic therapy, either as a vitamin K antagonist or aspirin. However, such current treatments are suboptimal, and despite recommendations, many patients do not receive adequate thromboprophylaxis. We discuss the reasons why guidelines are not adhered to in clinical practice and recent important advances.

Efficacy and Safety of Current Antithrombotic Therapy

Antithrombotic therapy is well established in the prevention of atrial fibrillation-related stroke.[5] Adjusted-dose warfarin reduced the risk of stroke by 64% compared with no antithrombotic treatment.[6] Aspirin alone reduced stroke by 22% compared with placebo.[6] Adjusted-dose warfarin reduced stroke risk by 38% versus aspirin.[6] The Atrial fibrillation Clopidogrel Trial with Irbesartan for prevention of Vascular Events (ACTIVE W) trial also has shown warfarin to be approximately 40% more efficacious than combined aspirin plus clopidogrel.[7,8] Many patients, however, receive aspirin because they are considered "unsuitable" for vitamin K antagonist therapy. In such patients, according to the ACTIVE A,[9] aspirin plus clopidogrel reduced the rate of major vascular events, in particular stroke, versus aspirin

Funding: None.

Conflict of Interest: Raffaele De Caterina is coauthor of the latest (2010) European Society of Cardiology Guidelines in Atrial Fibrillation and has participated in advisory boards related to atrial fibrillation therapies for Bayer, Bristol-Myers Squibb, Boehringer-Ingelheim, Daiichi Sankyo, Merck, Pfizer, and Sanofi-Aventis (overall: <$20,000). Elaine M. Hylek has participated in advisory boards related to atrial fibrillation therapies for Bayer, Bristol-Myers Squibb, Boehringer-Ingelheim, Daiichi Sankyo, Merck, Ortho-McNeil, and Pfizer (overall: <$10,000), and receives support from the National Institutes of Health (>$10,000).

Authorship: Both authors had access to the data and played a role in writing this manuscript.

Requests for reprints should be addressed to Raffaele De Caterina, MD, PhD, Institute of Cardiology and University Cardiology Division, Ospedale SS. Annunziata, Via dei Vestini, 66013 Chieti, Italy.

E-mail address: rdecater@unich.it

0002-9343/$ -see front matter
doi:10.1016/j.amjmed.2011.03.013

alone (relative risk 0.72; 95% confidence interval [CI], 0.62-0.84), but with an increased risk of major hemorrhage (relative risk 1.56; 95% CI, 1.28-1.89), and therefore with uncertain clinical benefit.[9]

Despite proven efficacy, vitamin K antagonists also cause more major bleeding. Compared with aspirin, vitamin K antagonists were associated with a 70% and 128% increase in the relative risk of major extracranial and intracranial hemorrhage, respectively,[6] a risk higher in the first 90 days of therapy among patients with the highest stroke risk.[10] Although contemporary risk of bleeding with better international normalized ratio control may be lower than before,[11] balancing the risk of stroke and bleeding among older individuals with several comorbidities is a continuing clinical challenge.[6]

CLINICAL SIGNIFICANCE

- Prevention of stroke in atrial fibrillation is an important part of atrial fibrillation management. Guidelines here mostly recommend a vitamin K antagonist, but are largely disattended.
- We discuss current gaps in knowledge, reasons why guidelines are not adhered to, and the need for new anticoagulants.
- We also highlight results so far obtained with novel anticoagulants in trials just completed.

Problems with Current Antithrombotic Therapies and Uncertainties with Current Risk Prediction Schemes

Guidelines for antithrombotic therapy in patients with atrial fibrillation are based on predicted stroke risk (Table 1). The most widely adopted risk stratification scheme has been $CHADS_2$,[12] based on a cumulative scoring system focusing on 5 major risk factors: **C**ongestive heart failure, **H**ypertension, **A**ge ≥75 years, and **D**iabetes, each scoring 1; and history of prior **S**troke or TIA, scoring 2 to reflect its increased weight.[12] A score of 0 indicates low risk, a score of 1 indicates moderate risk, and a score of 2 or more indicates high risk.[13] The definitions of low, moderate, and high risk of stroke used in the American College of Cardiology/American Heart Association/European Society of Cardiology (ACC/AHA/ESC) 2006 joint guidelines[14] and the American College of Chest Physicians guidelines[15] also are based on the presence or absence of these same risk factors in addition to impaired left ventricular function. However, because of variations in the criteria used by different risk stratification schemes, the proportion of patients categorized as low, moderate, or high risk may range from 7% to 42%.[16,17] In addition, less well-validated risk factors, such as female gender, age 65 to 74 years, and vascular disease, were not formally incorporated in older schemes. A new expanded version of $CHADS_2$ (CHA_2DS_2VASc) that includes these factors has recently been developed.[18] **A**ge ≥ 75 years is assigned 2 points (A_2), **a**ge 65 to 74 years is assigned 1 point (A), **v**ascular disease (prior myocardial infarction, peripheral arterial disease, complex aortic plaque) is assigned 1 point, and female gender (**s**ex) is assigned 1 point. Compared with existing schemes, it seems to have a slightly better predictive value for stroke and thromboembolism, with only 15% of patients in the moderate risk group. Furthermore, at 1 year no thromboembolic events were recorded for patients with a CHA_2DS_2VASc score of 0, thus probably allowing the identification of patients at "truly" low risk.[18] This scheme has been adopted by the newest ESC Guidelines as an add-on to the $CHADS_2$ score in patients in whom the latter is ≤ 1.[11]

Vitamin K Antagonists or Aspirin for Moderate Stroke Risk

The 2006 ACC/AHA/ESC guidelines recommended either a vitamin K antagonist or aspirin for individuals with a $CHADS_2$ score of 1,[14] in whom "risk of bleeding complications, ability to safely sustain adjusted chronic anticoagulation, and patients' preferences should be assessed."[14] This recommendation reflected concern for inducing harm from vitamin K antagonists among individuals with a relatively low risk of stroke, but assumed that aspirin is safer and that the choice of aspirin may be an appropriate balance between risk and benefit in such patients. Recent evidence, however, has shifted the focus toward anticoagulants for patients at moderate risk. The Birmingham Atrial Fibrillation Treatment of the Aged Study (BAFTA) randomized individuals aged more than 75 years, in whom the referring physicians needed to be uncertain as to which therapy provided the best benefit/risk balance, to either warfarin or aspirin. Fatal and nonfatal disabling stroke or systemic arterial embolism occurred at 1.8%/year for warfarin versus 3.8%/year for aspirin (hazard ratio [HR] 0.48; 95% CI, 0.28-0.80), with no difference in rates of major hemorrhage (1.9 vs 2.2, respectively).[19] In an analysis of ACTIVE W, reporting risks and benefits of vitamin K antagonists versus aspirin plus clopidogrel in relation to stroke risk ($CHADS_2$ score = 1 vs $CHADS_2$ >1)[20] and stroke rates for those with a $CHADS_2$ score of 1 were 1.25% and 0.43% per year on aspirin plus clopidogrel and oral vitamin K antagonists, respectively ($P = .01$). In patients with a score > 1, rates were 3.15% per year on aspirin plus clopidogrel and 2.01% on vitamin K antagonists ($P = .01$).[20] The rates of major bleeding for patients with a $CHADS_2$ score of 1 were comparable (1.36% on vitamin K antagonists vs 2.09% on aspirin plus clopidogrel, $P = .11$). The use of vitamin K antagonists gave a lower net risk (vascular events plus major bleeding) with vitamin K antagonists compared with aspirin plus clopidogrel (2.97% vs 5.25% per year; $P = .001$).[20] Therefore, as in 2 other recent trials,[21,22] patients with a $CHADS_2$ score of 1 apparently derive a modest (<1% per year) but significant benefit with vitamin K antagonists, with no in-

Table 1 2007 American College of Cardiology/American Heart Association/European Society of Cardiology, 2008 American College of Chest Physicians, and 2010 European Society of Cardiology Guidelines for the Use of Antithrombotic Therapy in Patients with Atrial Fibrillation

	ACC/AHA/ESC 1*		ACCP[2]		2010 ESC[13]	
Risk Category	Risk Factors	Recommended Therapy	Risk Factors	Recommended Therapy	Risk Factors	Recommended Therapy
Low	No risk	ASA 81-325 mg daily (class I)‡	Patients with AF, including paroxysmal AF, aged ≤ 75 y with no other risk factors	ASA 75-325 mg/d	No risk: CHA_2DS_2-VASc score = 0	ASA 75-325 mg/d or no antithrombotic therapy Preferred: no antithrombotic therapy rather than aspirin
Moderate	Patients with AF and 1 of the following moderate risk factors: age ≥ 75 y, hypertension, heart failure, LVEF ≤ 35%, diabetes mellitus	ASA 81-325 mg daily (class I)‡ or VKAs, target INR 2.5 (range 2.0-3.0) (class IIa)§	Patients with AF, including paroxysmal AF, with 1 of the following risk factors: age > 75 y, history of hypertension, diabetes mellitus, moderately or severely impaired LV systolic function, or heart failure	VKAs, target INR 2.5 (range 2.0-3.0), or ASA 75-325 mg/d Preferred OAC rather than ASA	CHA_2DS_2-VASc score = 1, ie, 1 "clinically relevant non-major" risk factor (heart failure or moderate to severe LV systolic dysfunction [eg, LVEF < 40%], hypertension, diabetes mellitus, female sex, age 65-74 y, vascular disease)	Either OAC‖ or ASA 75-325 mg/d Preferred: OAC rather than ASA
High	Patients with AF and prior thromboembolism (stroke, TIA, or systemic embolism), mitral stenosis,† prosthetic heart valve, or ≥ 2 of the above moderate risk factors	VKA, target INR 2.5¶ (range 2.0-3.0) (class I)‡	Patients with AF, including paroxysmal AF, and prior stroke, TIA, or systemic embolism or ≥ 2 of the above moderate risk factors	VKA, target INR 2.5 (range 2.0-3.0)	One or more "major" risk factor (previous stroke, TIA, or systemic embolism, age > 75 y) or ≥ 2 of the above "clinically relevant non-major" risk factors	OAC‖

ACC, American College of Cardiology; *ACCP,* American College of Chest Physicians; *AF,* atrial fibrillation; *AHA,* American Heart Association; *ASA,* acetylsalicylic acid; *ESC,* European Society of Cardiology; *INR,* international normalized ratio; *LV,* left ventricular; *LVEF,* left ventricular ejection fraction; *TIA,* transient ischemic attack; *VKA,* vitamin K antagonist. New OAC drugs, which may be viable alternatives to a VKAs, may ultimately be considered.

*Antithrombotic therapy is not recommended for patients with lone AF or contraindications.

†Refers to patients with valvular AF.

‡Treatment should be administered.

§It is reasonable to provide treatment.

‖OAC (eg, a VKA) adjusted to an intensity range of INR 2.0-3.0 (target 2.5).

¶If mechanical valve, target INR > 2.5.

crease in major hemorrhage.[20] Because of this, the 2010 ESC guidelines further restrict the uncertainty between vitamin K antagonists and aspirin indications only to patients with a CHA_2DS_2VASc score of 1 and, even in this case, with a preference for vitamin K antagonists.[11]

Underuse of Vitamin K Antagonists in Clinical Practice

Despite overwhelming evidence of the benefits of vitamin K antagonists on stroke risk in patients with atrial fibrillation, registries suggest that vitamin K antagonists are underused in clinical practice. In the National Anticoagulation Benchmark Outcomes Report, retrospectively evaluating practices in hospitalized patients with atrial fibrillation (n = 945), of 86% of patients eligible for warfarin, only 55% actually received it.[23] This proportion was similar for both academic and community hospitals. In the Euro Heart Survey on atrial fibrillation, only 67% of patients eligible for vitamin K antagonists were actually prescribed them, and 7% of eligible patients did not receive any form of antithrombotic treatment.[24] Three new large atrial fibrillation registries, the Global Anticoagulation Registry in the FIELD, the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation, and the RECORD-AF registry,[13,25] will shed further light on the patterns of use of vitamin K antagonists (and novel anticoagulants) worldwide.

Inherent Challenges of Vitamin K Antagonists

Some of the problems connected with the underuse of vitamin K antagonists are due to the intrinsic limitations of vitamin K antagonist. These include a narrow therapeutic window;[26,27] marked variability in dose-response, and the influence of environmental (eg, drug and food interactions) and genetic factors making it difficult for the patient to remain within the recommended optimal international normalized ratio of 2.0 to 3.0, necessitating regular monitoring and dose adjustments.[28,29] The management of vitamin K antagonists is further complicated by their delayed onset of action, with a complete effect not achieved until functional vitamin K-dependent coagulation factors (F II, VII, IX, and X) are cleared,[29] thus sometimes requiring bridging therapy with a parenteral anticoagulant.[29] Vitamin K antagonists also have a slow offset of action, related to their long half-lives (36-42 hours for warfarin) and the time required for the synthesis of new, functional coagulation factors.[28]

A number of physician-related factors also influence the use of vitamin K antagonists.[30,31] These include a perception of the presence of a potential contraindication, their underrating, a perceived risk of bleeding, and low patient compliance.[30] There is a continuing reluctance among physicians to prescribe vitamin K antagonists to the elderly, particularly patients aged >80 years,[32] in whom hypertension, renal impairment, and diabetes are known to increase bleeding with vitamin K antagonist therapy.[31] Cognitive impairment also decreases time in the therapeutic range, which is a good measure of anticoagulation quality,[33] thereby correctly increasing physicians' concerns.[31] Although newer guidelines recommend the adoption of bleeding risk stratification schemes, such as the HAS-BLED score (a score by which **H**ypertension (**H**) is given 1 point, **A**bnormal renal and liver functions (**A**) are given 1 point each, **S**troke (**S**) is given 1 point, **B**leeding (**B**) is given 1 point, **L**abile INRs (**L**) is given 1 point, **E**lderly (**E**, age >65 years) is given 1 point, **D**rug or alcohol abuse (**D**) are given 1 point each, up to a maximum of 9 points),[11] it is uncertain how these will separate the bleeding risk from the thromboembolic risk.

Novel Oral Anticoagulants: Latest Developments

The limitations of vitamin K antagonists have highlighted the need for new oral anticoagulants that may overcome their drawbacks while maintaining or improving therapeutic benefit. Ideally, such drugs would have predictable pharmacokinetic and pharmacodynamic profiles; limited or no drug/food interactions; a fast onset of action, removing the need for bridging therapy; and a fast offset of action to allow temporary discontinuation of anticoagulant therapy if required (eg, for surgery).[34] They also would have the potential to be given at a fixed dose with no monitoring or dose adjustment, thus being more convenient for patients and physicians.[34] Such anticoagulants might overcome the barriers to vitamin K antagonist use, leading to improved patient management.

Established vitamin K antagonists act on multiple targets of the coagulation cascade, with different and variable half-life. To make the drug effects more predictable, efforts have focused primarily on the direct inhibition of a single coagulation factor, in particular FIIa (thrombin) and FXa.[35] The potential of novel oral anticoagulants for stroke prevention in atrial fibrillation was first demonstrated with the direct thrombin inhibitor ximelagatran in the stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation (SPORTIF) trials.[36-38] However, its use was associated with liver toxicity,[39] which led to the termination of its clinical development.[40]

Studies on several new oral anticoagulants have been recently completed or are currently undergoing phase III clinical trials in patients with atrial fibrillation. These studies aim at demonstrating non-inferiority compared with vitamin K antagonists or superiority compared with aspirin in patients for whom vitamin K antagonists are contraindicated or not tolerated.

Oral Direct Thrombin Inhibitors

Dabigatran etexilate. The recently published phase III Randomized Evaluation of Long-term anticoagulation therapY (RE-LY) study[41] was a prospective, multicenter, parallel-group, non-inferiority trial evaluating the efficacy and safety of 2 fixed doses of dabigatran etexilate (110 mg and 150 mg twice per day [bid]), in a blinded fashion, compared with open-label

warfarin in patients with atrial fibrillation at increased risk of stroke. Dabigatran 150 mg bid was superior to warfarin for the primary efficacy outcome of stroke and systemic embolism (1.11% vs 1.69% per year, $P < .001$), with a similar rate of major bleeding (3.11% vs 3.36% per year, $P = .31$), whereas dabigatran 110 mg bid was non-inferior to warfarin for the primary efficacy outcome (1.53% vs 1.69% per year, $P = .34$), with a significantly lower rate of major bleeding (2.71% vs 3.36% per year, $P = .003$).[41] Both dabigatran doses also had lower rates of intracranial hemorrhage, including hemorrhagic stroke and subdural or subarachnoid hemorrhage,[41] and no evidence of hepatotoxicity. A higher rate of myocardial infarction was observed with both doses of dabigatran, but neither was significant compared with warfarin (0.82 for the 110 mg dose and 0.81 for the 150 mg dose vs 0.64 for warfarin; $P = .09$ and 0.12, respectively). Overall vascular mortality was reduced with dabigatran.[41] Dyspepsia was reported by approximately 11.5% of participants assigned to either dose of dabigatran versus 5.8% for warfarin ($P < .001$).

Dabigatran is now the first oral anticoagulant with the potential to replace vitamin K antagonists for the long-term prevention of thromboembolism in patients with atrial fibrillation, having been recently approved by the US Food and Drug Administration at the dosage of 150 mg bid and by the Canadian regulatory authorities at both dosages tested in the RE-LY study. Further safety analysis of dabigatran will be carried out in the RELY-ABLE trial (NCT00808067, estimated completion date: July 2011).

Oral Direct Factor Xa Inhibitors

Rivaroxaban. Rivaroxaban has been assessed in the phase III Rivaroxaban Once daily oral direct FXa inhibition Compared with vitamin K antagonism for prevention of stroke and Embolism Trial in Atrial Fibrillation (ROCKET-AF) (NCT00403767). This was a randomized, double-blind, double-dummy, multicenter, event-driven study to test the non-inferiority of rivaroxaban (20 mg/d or 15 mg/d in patients with renal impairment [creatinine clearance of 30-49 mL/min]) compared with adjusted-dose warfarin in subjects with non-valvular atrial fibrillation at high risk of stroke. The primary safety outcome was the composite of major and non-major clinically relevant bleeds. As reported at the 2010 American Heart Association Scientific Sessions (http://www.theheart.org/article/1148785.do), the trial met the non-inferiority end point, with overall rates of bleeding for rivaroxaban similar to those for warfarin, but with less intracranial hemorrhage. A comparison of the characteristics and main outcomes of RELY and ROCKET-AF is shown in Table 2.

Apixaban. Apixaban for the prevention of stroke in subjects with atrial fibrillation (ARISTOTLE, NCT00412984) is a phase III randomized, double-blind, double-dummy study testing the non-inferiority of apixaban versus warfarin for the composite of stroke and systemic embolism in approximately 18,000 patients with atrial fibrillation. Patients are required to have at least 1 additional risk factor for stroke, including age ≥ 75 years, previous stroke, transient ischemic attack or systemic embolism, and diabetes. The study will be complete in April 2011.

Apixaban has already been assessed in the phase III Study of Apixaban in Patients with Atrial Fibrillation (AVERROES, NCT00496769), a randomized, double-blind, double-dummy study to assess the superiority of apixaban 5 mg bid versus aspirin (81-324 mg/d) for the prevention of stroke in 5600 patients with atrial fibrillation and at least 1 additional risk factor for stroke, who had failed or were considered unsuitable for vitamin K antagonist treatment for reasons including poor anticoagulation control, adverse events, and the need for other treatments that may interact with vitamin K antagonists. The primary efficacy outcome was the time to first ischemic stroke, hemorrhagic stroke, or systemic embolism. In April 2010, the Data and Safety Monitoring Board recommended early study termination because of clear benefit in favor of apixaban. The median duration of follow-up was 1.5 years. The primary outcome was stroke or systemic embolism. There were 52 primary outcome events in those randomized to apixaban (1.6%/year) and 112 primary outcome events in those randomized to aspirin (3.5%/year) (HR 0.46; 95% CI, 0.33-0.64; $P < .001$). Mortality rates were 3.4%/year for those randomized to apixaban and 4.4%/year for those randomized to aspirin (HR 0.79; 95% CI, 0.61-1.01; $P = .06$). There were 46 major bleeds (1.4%/year) in the apixaban group and 43 major bleeds (1.3%/year) in the aspirin group (HR 1.08; 95% CI, 0.71-1.63; $P = .73$). There were 13 intracranial bleeds in the apixaban group and 12 intracranial bleeds in the aspirin group.[42] Thus, a clear superiority of apixaban over aspirin was shown in terms of efficacy, with comparable safety.

Edoxaban (DU-176b). The Global Study to Assess the Safety and Effectiveness of DU-176b vs Standard Practice of Dosing with Warfarin in Patients with Atrial Fibrillation (ENGAGE AF-TIMI 48, NCT00781391) is a randomized double-blind, double-dummy study in subjects with atrial fibrillation and a high risk of stroke ($CHADS_2$ score ≥ 2). The study will evaluate the efficacy and safety of 2 doses of edoxaban (30 or 60 mg/d, with halved doses for moderate renal impairment or low body weight) compared with warfarin. The primary study hypothesis is that at least 1 dose regimen of edoxaban will be non-inferior to warfarin in reducing the composite of stroke and systemic embolism. The trial is estimated to be completed in March 2012.

Other agents, such as the direct thrombin inhibitor AZD0837 (13) and the direct FXa inhibitors betrixaban (NCT00742859) and YM-150 (NCT00448214), are in phase II clinical development in the atrial fibrillation setting.

Other Oral Anticoagulants Under Development

Two other classes of oral anticoagulants are now in earlier stages of development. Tecarfarin, now in phase 2 devel-

Table 2 Comparison of RE-LY and ROCKET-AF Results*

	RE-LY[46]		ROCKET AF*
Comparisons	Dabigatran 110 mg bid vs warfarin	Dabigatran 150 mg bid vs warfarin	Rivaroxaban 20 mg qd vs warfarin
Study design	Dabigatran doses assigned in a double-blind fashion; warfarin given open-label; blinded adjudication of outcomes		double-blind, double-dummy
Sample size	18,113 patients, 3 arms (dabigatran 110 mg bid, dabigatran 150 mg bid, or dose-adjusted warfarin (titrated to an INR of 2.5).		14,000 patients, 2 arms: 20 mg qd (or 15 mg in patients with moderate renal impairment at screening) or to dose-adjusted warfarin (titrated to an INR of 2.5).
Quality of warfarin comparison (% TTR)	64 (mean)		58 (median)
Mean age (y)	71		73
Mean $CHADS_2$ score	2.1		3.5
Prior use of VKAs (%)	50%		63%
Primary outcome	Stroke and systemic embolism		Stroke and systemic embolism
Primary end point, HR (95% CI), non-inferiority, *P* value*	0.91 (0.74-1.11) <.001	0.66 (0.53-0.82), <.001	0.79 (0.66-0.96) <.001
Primary end point, HR (95% CI), on-treatment superiority, *P* value			0.79 (0.65-0.95) .015
Primary end point, HR (95% CI), intention-to-treat superiority, *P* value	0.91 (0.74-1.11) .34	0.66 (0.53-0.82) <.001	0.88 (0.74-1.03) .117
Hemorrhagic stroke, HR (95% CI) *P* value	0.31 (0.17-0.56) <.001	0.26 (0.14-0.49) <.001	0.59 (0.37-0.93) .024
Major and minor bleeding, HR (95% CI), *P* value	0.78 (0.74-0.83) <.001	0.91 (0.86-0.97) .002	1.03 (0.96-1.11) .442

bid, Twice per day; *CI,* confidence interval; *HR,* hazard ratio; *INR,* international normalized ratio; *qd,* once per day; *TTR,* time in the therapeutic range; *VKA,* vitamin K antagonist.

*Based on the limited results available for ROCKET-AF from the 2010 AHA Scientific Sessions Presentation. For ROCKET AF, the non-inferiority analysis shown was in the on-treatment population.

opment, is a vitamin K antagonist metabolized through the esterase pathways,[43] overcoming the limitations related to warfarin metabolism through cytochrome P450, namely, the many drug-drug interactions.

TTP889 is a selective, small-molecule, orally available, partial FIXa inhibitor, which can be administered with fixed once per day dosing, with a reported half-life of 21 to 25 hours.[44] Preclinical data support the concept of upstream coagulation inhibition of FIXa (more upstream than FXa) as a viable and possibly highly efficient way to prevent thrombosis.

CONCLUSIONS

The choice of antithrombotic therapy for patients with atrial fibrillation currently depends on the appropriate classification of an individual patient's predicted risk of stroke. Most recent guidelines overall recommend oral anticoagulants (vitamin K antagonists) for most patients with atrial fibrillation, even at moderate risk of stroke. On the other hand, there is a clear underuse of vitamin K antagonists and a strong need for safer and more convenient antithrombotic drugs. The new oral anticoagulants in development, with fewer food and drug interactions and predictable pharmacology, have no requirement for routine coagulation monitoring and may provide equally or more effective, yet safer, alternative options.

ACKNOWLEDGMENTS

The authors thank Rachel Bell and Valentina Berti for expert editorial assistance.

References

1. Heeringa J, van der Kuip DA, Hofman A, et al. Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. *Eur Heart J.* 2006;27:949-953.
2. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke.* 1991;22:983-988.
3. Jorgensen HS, Nakayama H, Reith J, et al. Acute stroke with atrial fibrillation. The Copenhagen Stroke Study. *Stroke.* 1996;27:1765-1769.
4. Hughes M, Lip GY. Stroke and thromboembolism in atrial fibrillation: a systematic review of stroke risk factors, risk stratification schema and cost effectiveness data. *Thromb Haemost.* 2008;99:295-304.

5. De Caterina R, Husted S, Wallentin L, et al. Anticoagulants in heart disease: current status and perspectives. *Eur Heart J.* 2007;28:880-913.
6. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Ann Intern Med.* 2007;146:857-867.
7. Connolly S, Pogue J, Hart R, et al. Clopidogrel plus aspirin vs oral anticoagulation for atrial fibrillation in the Atrial fibrillation Clopidogrel Trial with Irbesartan for prevention of Vascular Events (ACTIVE W): a randomised controlled trial. *Lancet.* 2006;367:1903-1912.
8. Connolly SJ, Pogue J, Eikelboom J, et al. Benefit of oral anticoagulant over antiplatelet therapy in atrial fibrillation depends on the quality of international normalized ratio control achieved by centers and countries as measured by time in therapeutic range. *Circulation.* 2008;118:2029-2037.
9. Connolly SJ, Pogue J, Hart RG, et al. Effect of clopidogrel added to aspirin in patients with atrial fibrillation. *N Engl J Med.* 2009;360:2066-2078.
10. Hylek EM, Evans-Molina C, Shea C, et al. Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. *Circulation.* 2007;115:2689-2696.
11. Camm AJ, Kirchhof P, Lip GY, et al. Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *Eur Heart J.* 2010;31:2369-2429.
12. Gage BF, Waterman AD, Shannon W, et al. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA.* 2001;285:2864-2870.
13. Lip GY, Rasmussen LH, Olsson SB, et al. Oral direct thrombin inhibitor AZD0837 for the prevention of stroke and systemic embolism in patients with non-valvular atrial fibrillation: a randomized dose-guiding, safety, and tolerability study of four doses of AZD0837 vs. vitamin K antagonists. *Eur Heart J.* 2009;30:2897-2907.
14. Fuster V, Ryden LE, Cannom DS, et al. ACC/AHA/ESC 2006 Guidelines for the Management of Patients with Atrial Fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Revise the 2001 Guidelines for the Management of Patients With Atrial Fibrillation): developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. *Circulation.* 2006;114:e257-354.
15. Singer DE, Albers GW, Dalen JE, et al. Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest.* 2008;133(6 Suppl):546S-592S.
16. Stroke Risk in Atrial Fibrillation Working Group. Comparison of 12 risk stratification schemes to predict stroke in patients with nonvalvular atrial fibrillation. *Stroke.* 2008;39:1901-1910.
17. Fang MC, Go AS, Chang Y, et al. Comparison of risk stratification schemes to predict thromboembolism in people with nonvalvular atrial fibrillation. *J Am Coll Cardiol.* 2008;51:810-815.
18. Lip GY, Nieuwlaat R, Pisters R, et al. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the Euro Heart Survey on Atrial Fibrillation. *Chest.* 2010;137:263-272.
19. Mant J, Hobbs FD, Fletcher K, et al. Warfarin vs aspirin for stroke prevention in an elderly community population with atrial fibrillation (the Birmingham Atrial Fibrillation Treatment of the Aged Study, BAFTA): a randomised controlled trial. *Lancet.* 2007;370:493-503.
20. Healey JS, Hart RG, Pogue J, et al. Risks and benefits of oral anticoagulation compared with clopidogrel plus aspirin in patients with atrial fibrillation according to stroke risk: the atrial fibrillation clopidogrel trial with irbesartan for prevention of vascular events (ACTIVE-W). *Stroke.* 2008;39:1482-1486.
21. Lee BH, Park JS, Park JH, et al. The effect and safety of the antithrombotic therapies in patients with atrial fibrillation and CHADS score 1. *J Cardiovasc Electrophysiol.* 2010;21:501-507.
22. Gorin L, Fauchier L, Nonin E, et al. Antithrombotic treatment and the risk of death and stroke in patients with atrial fibrillation and a CHADS2 score=1. *Thromb Haemost.* 2010;103:833-840.
23. Waldo AL, Becker RC, Tapson VF, Colgan KJ. Hospitalized patients with atrial fibrillation and a high risk of stroke are not being provided with adequate anticoagulation. *J Am Coll Cardiol.* 2005;46:1729-1736.
24. Nieuwlaat R, Capucci A, Camm AJ, et al. Atrial fibrillation management: a prospective survey in ESC member countries: the Euro Heart Survey on Atrial Fibrillation. *Eur Heart J.* 2005;26:2422-2434.
25. Le Heuzey JY, Breithardt G, Camm J. Profile of patients according to chosen therapeutic strategy in the RecordAF global registry on atrial fibrillation management. *Circulation.* 2008;118(Abs 4634):S923.
26. Hylek EM, Skates SJ, Sheehan MA, Singer DE. An analysis of the lowest effective intensity of prophylactic anticoagulation for patients with nonrheumatic atrial fibrillation. *N Engl J Med.* 1996;335:540-546.
27. Hirsh J, Fuster V, Ansell J, Halperin JL. American Heart Association/American College of Cardiology Foundation guide to warfarin therapy. *J Am Coll Cardiol.* 2003;41:1633-1652.
28. Ansell J, Hirsh J, Hylek E, et al. Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest.* 2008;133(6 Suppl):160S-198S.
29. Lin PJ. Reviewing the reality: why we need to change. *Eur Heart J Suppl.* 2005;:E15-E20.
30. Deplanque D, Leys D, Parnetti L, et al. Stroke prevention and atrial fibrillation: reasons leading to an inappropriate management. Main results of the SAFE II study. *Br J Clin Pharmacol.* 2004;57:798-806.
31. Robert-Ebadi H, Le Gal G, Righini M. Use of anticoagulants in elderly patients: practical recommendations. *Clin Interv Aging.* 2009;4:165-177.
32. Friberg L, Hammar N, Ringh M, et al. Stroke prophylaxis in atrial fibrillation: who gets it and who does not? Report from the Stockholm Cohort-study on Atrial Fibrillation (SCAF-study). *Eur Heart J.* 2006;27:1954-1964.
33. Flaker GC, Pogue J, Yusuf S, et al. Cognitive function and anticoagulation control in patients with atrial fibrillation. *Circ Cardiovasc Qual Outcomes.* 2010;3:277-283.
34. Halperin JL. What can ongoing clinical trials of anticoagulants demonstrate? *J Cardiovasc Med (Hagerstown).* 2009;10:610-615.
35. Lassen MR, Laux V. Emergence of new oral antithrombotics: a critical appraisal of their clinical potential. *Vasc Health Risk Manag.* 2008;4:1373-1386.
36. Olsson SB. Stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation (SPORTIF III): randomised controlled trial. *Lancet.* 2003;362:1691-1698.
37. Albers GW, Diener HC, Frison L, et al. Ximelagatran vs warfarin for stroke prevention in patients with nonvalvular atrial fibrillation: a randomized trial. *JAMA.* 2005;293:690-698.
38. Diener HC. Stroke prevention using the oral direct thrombin inhibitor ximelagatran in patients with non-valvular atrial fibrillation. Pooled analysis from the SPORTIF III and V studies. *Cerebrovasc Dis.* 2006;21:279-293.
39. Lee WM, Larrey D, Olsson R, et al. Hepatic findings in long-term clinical trials of ximelagatran. *Drug Saf.* 2005;28:351-370.
40. AstraZeneca. AstraZeneca decides to withdraw Exanta™. Available at: http://www.astrazeneca.com/Media/Press-releases/Article/20060214—AstraZeneca-Decides-to-Withdraw-Exanta.
41. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran vs warfarin in patients with atrial fibrillation. *N Engl J Med.* 2009;361:1139-1151.
42. Connolly SJ, Eikelboom J, Joyner C, et al. Apixaban in patients with atrial fibrillation. *N Engl J Med.* 2011;364:806-817.
43. Ellis DJ, Usman MH, Milner PG, et al. The first evaluation of a novel vitamin K antagonist, tecarfarin (ATI-5923), in patients with atrial fibrillation. *Circulation.* 2009;120:1029-1035, 2 p following 1035.
44. Eikelboom JW, O'Donnell M, Yusuf S, et al. Rationale and design of AVERROES: apixaban vs acetylsalicylic acid to prevent stroke in atrial fibrillation patients who have failed or are unsuitable for vitamin K antagonist treatment. *Am Heart J.* 2010;159:348-353 e1.

Use of statins and recurrence of atrial fibrillation after catheter ablation or electrical cardioversion

A systematic review and meta-analysis

Francesco Dentali[1]; Monica Gianni[2]; Alessandro Squizzato[1]; Walter Ageno[1]; Luana Castiglioni[1]; Lorenzo Maroni[1]; Elaine M. Hylek[3]; Anna Maria Grandi[1]; Eugenio Cazzani[2]; Achille Venco[1]; Luigina Guasti[1]

[1]Department of Clinical Medicine, University of Insubria, Varese, Italy; [2]Department of Cardiology, Hospital of Tradate, Varese, Italy; [3]Department of Medicine, Research Unit – Section of General Internal Medicine, Boston University Medical Center, Boston, Massachusetts, USA

Summary

Statins have important pleiotropic effects and have been shown to reduce vascular inflammation. Some evidence suggests that statins may have a role in the primary prevention of atrial fibrillation (AF), whereas little is know on the role of statins in patients with existing AF. We performed a meta-analysis of the literature to assess the effect of statins on the recurrence of AF after electrical cardioversion or ablation. MEDLINE and EMBASE databases were searched up to January 2010. Relative risks (RR) and 95% confidence intervals (CIs) were then calculated and pooled using a random-effects model. Statistical heterogeneity was evaluated through the use of I^2 statistics. Sixteen studies were included in our systematic review. Statins did not reduce the risk of AF recurrence after ablation (four studies including 750 patients; RR, 1.04; 95% CI, 0.85–1.28, p=0.71; I^2 = 34%). Conversely, the use of statins was associated with a significantly reduced risk of AF recurrence after electrical cardioversion (12 studies including 1790 patients; RR, 0.78; 95% CI, 0.67–0.90, p=0.0003; I^2 = 34%). This reduction was not statistically significant when the analysis was restricted to randomised controlled trials (RCTs) only (five studies, 458 patients, RR, 0.76; 95% CI, 0.48–1.20). In conclusion, statins may lower the risk of AF recurrence after electrical cardioversion, but not ablation. However, this finding should be considered with caution, and larger RCTs are warranted to confirm our preliminary results.

Keywords

Statins, atrial fibrillation

Correspondence to:
Francesco Dentali
U.O. Medicina Interna, Ospedale di Circolo
Viale Borri 57, 21100 Varese, Italy
Tel.: +39 0332 278594, Fax: +39 0332 278229
E-mail: fdentali@libero.it

Received: October 17, 2010
Accepted after major revision: April 29, 2011
Prepublished online: May 26, 2011
doi:10.1160/TH10-10-0660
Thromb Haemost 2011; 106: 363–370

Introduction

AF is the most common arrhythmia encountered in clinical practice with an increasing prevalence during the last few decades (1). Prevalence of AF is increasing with age (1, 2) and these patients have a higher prevalence of hypertension and a larger waist circumference than non- AF patients (2). On the other hand, presence of hypercholesterolaemia seems to be less common in AF patients (2, 3). AF is associated with a significant increase in the risk of cardiovascular morbidity and cardiovascular overall mortality (4) representing a major public health problem (1). Although control of the ventricular response is an acceptable treatment in certain subgroups of patients, restoration and maintenance of sinus rhythm remains a widely used strategy in many patients (4). This strategy offers several potential benefits, such as the prevention of electrical and structural remodelling of the atria, improved haemodynamic function, amelioration of symptoms, and improvement in the quality of life (4).

Although some recent studies indicate increased efficacy of pharmacological cardioversion using combination therapy (5), electrical cardioversion is the most commonly used method for sinus rhythm restoration in patients with persistent AF. Despite the use of antiarrhythmic agents for sinus rhythm maintenance, a considerable proportion of patients relapse to AF (1, 6–9).

Catheter ablation has been proposed as an effective therapeutic option for AF that is resistant to pharmacologic rhythm or rate control, with successful long-term maintenance of sinus rhythm in the absence of treatment with anti-arrhythmic drugs reported in many patients (10). However, the recurrence rate of AF after catheter ablation has been reported to range between 30 and 40%, depending on the ablation strategy and the type of AF (10, 11).

Factors associated with relapse include older age, atrial dilation, and longer duration of the arrhythmia. Recently, experimental and clinical studies have demonstrated an association between AF and inflammation, suggesting a role of inflammation both in the genesis and in the recurrence of AF (12–14). A number of studies and a meta-analysis indicate a positive association between C-reactive protein (CRP) levels and AF recurrence (15).

Hydroxymethylglutaryl coenzyme-A reductase inhibitors (statins) have anti-oxidant effects (16) and they decrease inflammatory markers independent of their action on lipids (17). Both experimental and clinical studies have demonstrated that statins pre-

vent remodelling and reduce the incidence of AF (18). Clinical trials and a recently published meta-analysis of the literature have suggested that statins reduce the incidence of AF after cardiac surgery (19). On the other hand, evidence on the role of statins in preventing AF recurrence after electrical cardioversion or ablation is less compelling (20).

Therefore we performed a systematic review and a meta-analysis of the literature to assess the effect of statins on recurrence of AF after electrical cardioversion or ablation.

Methods

A protocol was prospectively developed, detailing the specific objectives, criteria for study selection, approach to assess study quality, outcomes, and statistical methods.

This systematic review was performed according to the guidelines for Quality of Reporting of Meta-analysis (PRISMA, MOOSE) (21, 22).

Study identification

We tried to identify all published studies that evaluated the role of statins on recurrence of AF after electrical cardioversion or ablation using the MEDLINE (1966 to January 2010, week 4) and EMBASE (1980 to January 2010, week 4) electronic databases. The search strategy was developed without any language restriction, and used the subject headings and key words presented in Appendix 1 (see supplementary material online available at www.thrombosis-online.com). We supplemented our search by manually reviewing the reference lists of all retrieved articles for additional published or unpublished trials and by searching the abstracts of the American Heart Association (AHA) (from 1999 to 2009) and the European Society of Cardiology (ESC) (from 2005 to 2009) Scientific Meetings. Abstracts presented at the ESC Scientific Meetings and at AHA Scientific Sessions were searched at http://spo.escardio.org/abstract-book/search.aspx and at www.ahajournals.org, respectively.

Study selection

Study selection was performed independently by two reviewers (FD, MG), with disagreements resolved through discussion and by the opinion of a third reviewer (LG), if necessary. Studies were included if they met the following criteria: 1) separate data for patients on treatment with statins and controls were available; 2) recurrence of AF was objectively documented. Both observational and experimental studies were included.

Studies not including a control group drawn from the same population, animal studies, *in vitro* studies, or trials that exclusively reported other clinical outcomes were excluded. Case reports, editorials, commentaries, letters, review articles, guidelines or secondary prevention trials were also excluded from the analysis.

To assess the agreement between reviewers for study selection, we used the kappa (κ) statistic, which measures agreement beyond chance (23).

When more than one publication from the same patient cohort existed, then the study with the most complete data set was included in the systematic review.

Data extraction

Two reviewers (FD, MG) independently extracted data on study (year of publication, design, study centre) and patients characteristics (number of subjects enrolled, mean age, variation in age, gender and race). Furthermore, the following characteristics were collected: (i) total follow-up duration for randomised controlled trials (RCTs) and cohort studies; (ii) total number of patients with AF recurrence (iii) molecule and regimen of statins.

In case necessary data were not provided in the manuscript, the corresponding author was contacted for additional data request.

Study validity assessment

Two unmasked investigators (FD, MG) independently completed the assessment of study validity. For RCTs, we planned quality assessment by means of Jadad's scale, which evaluates the following three study characteristics: method of randomisation, method of blinding, and follow-up (24). To stratify RCTs, we applied the following cut-offs: a total of five points defined high quality studies; three and four points defined medium quality studies; two or less points defined low quality studies.

Although in observational studies the use of quality scoring systems or quality scales is controversial (22), study quality was assessed by the following items for cohort studies: type of study (prospective or retrospective); patient selection (consecutive patients without potential bias of selection); control group (consecutive enrolment or matched for age and sex). For each fulfilled item one point was given. A scoring system was adapted to identify two quality categories as follows: a total of three points defined high quality studies; two or less points defined low quality studies. The total number of patients lost to follow-up (less than 5% of patients, more than 20%, or between 5 and 20%) was also ascertained as an additional quality item.

Statistical analysis

Relative risk (RR) and 95% confidence intervals (CIs) of AF recurrence after catheter ablation and electrical cardioversion were cal-

culated. The data were pooled using a random-effects model (the DerSimonan and Laird method) (25). Separate analyses for patients undergoing catheter ablation and electrical cardioversion were performed. For treatment effects that were statistically significant, we determined the absolute risk reduction and NNT to prevent a recurrence. Statistical heterogeneity was evaluated using the I^2 statistic, which assesses the appropriateness of pooling the individual study results (26). The I^2 value provides an estimate of the amount of variance across studies due to heterogeneity rather than chance. $I^2 < 30\%$ indicates mild heterogeneity, 30–50% moderate, and > 50% severe heterogeneity. When heterogeneity was present, we repeated the analysis removing one study at a time to assess the source of heterogeneity. Presence of publication bias was explored using funnel plots of effect size against standard error (27). The software Review Manager (RevMan, version 5.0.16 for Windows, Oxford, UK; The Cochrane Collaboration, 2008) supported the analysis.

As a sensitivity analysis, we planned to analyze separately RCTs considering the effect of treatment with statins.

Results

Study identification and selection

We identified 1,475 potentially relevant studies from the following databases: 1,356 from EMBASE and 119 from MEDLINE. Further 673 abstracts from the American Heart Association and European Society of Cardiology Scientific Meeting Abstracts were found using "atrial fibrillation" and "statins" search terms. We excluded 2,094 studies after title and abstract screening using the predefined inclusion and exclusion criteria; the remaining 54 studies were retrieved in full for detailed evaluation. Two additional studies were identified through manual review of references. Agreement between reviewers for study selection was optimal (K=0.91). Of the 56 retrieved studies, 40 were excluded for the following reasons: 27 did not match inclusion criteria, 10 were editorials or commentaries, two reported duplicate data, and one RCT could not be included since the authors did not provide the absolute number of AF recurrences (28). Sixteen studies (29–44) were therefore included in this systematic review: 15 were published as full text and one as an abstract. The study identification and selection progression is detailed in Appendix 2 (see supplementary material online available at www.thrombosis-online.com).

Study characteristics

Baseline characteristics of patients included in the studies are summarised in ▶Table 1. Five RCTs, nine retrospective cohort studies, and two prospective cohort studies were included in our systematic review. Study size ranged from 44 to 625 patients, for a total of 2,540 patients. AF recurrence was the primary end point of all included studies.

Study quality

Quality assessment items are summarized in Appendix 3 (see supplementary material online available at www.thrombosis-online.com). One of the five RCTs was of high quality. All 11 cohort studies were of low quality.

Catheter ablation

Recurrence of AF after catheter ablation was evaluated in four retrospective cohort studies for a total of 747 patients. Follow up periods varied from 30 days to 18 months. Type and dose of statins used in different studies were not specified. Furthermore, no study indicated when statins were started. AF recurrence occurred in 136 of 297 (45.8%) patients on treatment with statins and in 178 of 450 (39.6%) patients not on treatment with statins. The use of statins was not associated with a reduced risk of AF recurrence (RR 1.04, 95%CI 0.85, 1.28; p= 0.71) (▶Fig. 1). Heterogeneity across the studies was low (I^2= 34%).

Due to the low number of studies, funnel-plot analysis could not be done. Therefore, the presence of publication bias could not be excluded.

Electrical cardioversion

Recurrence of AF after electrical cardioversion was evaluated in 12 studies (5 RCTs, 5 retrospective cohort studies, and 2 prospective cohort studies) for a total of 1,790 patients. Statin regimen was not specified in five of the eight observational studies and the other three used seven different types of statins. Furthermore, none of the observational studies indicated when statins were started. Atorvastatin (at a dose varying from 10 to 80 mg) was used in three RCTs, pravastatin (dose of 40 mg) in one, and rosuvastatin (dose of 10 mg) in one. In these studies, statin initiation varied from three weeks to 48 hours before the electrical cardioversion. Follow up intervals varied from 30 days to more than 3.5 years. AF recurrence occurred in 179 of 475 (38.5%) patients on treatment with statins and in 606 of 1,325 (45.7%) patients not on treatment with statins. The use of statins resulted in a statistically significant reduction in the risk of AF recurrence (RR 0.78, 95%CI 0.67, 0.90; p< 0.001) (▶Fig. 1). Heterogeneity across the studies was low (I^2= 17%). The absolute risk reduction was 7.2% with a NNT of 14.

Funnel plot of RR versus standard error appeared slightly asymmetric with an absence of studies in the bottom right hand corner (see Appendix 4; supplementary material online available at www.thrombosis-online.com).

Sensitivity analysis included five RCTs for a total of 458 patients. AF recurrence occurred in 86 of 229 (37.6%) patients on treatment with statins and in 108 of 229 (47.2%) patients not on treatment with statins. The use of statins was associated with a statistically non significant reduction in the risk of AF recurrence in this subgroup of studies (RR 0.76, 95%CI 0.48, 1.20).

Table 1: Baseline characteristics of included studies.

Author	Population	Number of patients	Type of statin and dosage	Follow-up	Exclusion criteria
Almroth, 2009	Persistent AF undergoing EC	234 (65 ± 10 years)	Atorvastatin 80 mg die (14 days before EC and 30 days after)	30 days	• Age < 18 and > 80 years • Paroxysmal AF or atrial flutter • Contraindication against atorvastatin • Ongoing treatment with lipid-lowering drugs • Ongoing treatment with class I or class III antiarrhythmic • Oral amiodarone < 6 months • Known liver disease or a myopathy • Previous EC = 1 year
Xia, 2009	Persisten AF (> 48 hours) undergoing EC	64	Rosuvastatin 10 mg die (48 hours before and 3 months after)	3 months	• Age < 18 and > 75 years • Paroxysmal AF • Left atrium size > 55 mm • Left atrium trombi • Contraindication to statin treatment • Already in statin treatment • Hypertension • Diabetes mellitus • RF • CAD • History of smoking • Thyroid disfunction • Known rheumatic disease or cancer • Infection < 2 months
Can, 2007	Persistent AF undergoing EC	44	Atorvastatin 40 mg (started 3 weeks before EC)	2 m	–
Ozaydin, 2006	Persisten AF undergoing EC	48 (62 ± 11 years)	Atorvastitin 10 mg (48 hours before and 3 months after)	3 months	• Paroxysmal AF • Left atrium size > 6.5 cm • Moderate to severe heart valve disease • CAD • Unsuccessful EC • NYHA class III or IV heart failure • Cardiac surgery history • Acute reversible condition • Pregnancy or lactation • Contraindication to statin treatment • Already in statin treatment • Hyperthyroidism • Age < 18 years • EF < 30% • Significant RF • Low ejection fraction
Tveit, 2004	AF > 48 hours undergoing EC	114	Pravastatin 40 mg (3 weeks before and 6 weeks after)	6 weeks	• Significant heart valve disease • Cardiothoracic surgery in previous 30 days • Hyperthyroidism • Liver disease • Pregnancy or lactation • Already in statin treatment
Naji, 2009	Persistent AF undergoing EC	198	NS	2 years	• Duration of AF less than one months • Age >85 years • Heart surgery or electrophysiologic procedure prior EC or during follow-up • Implanted pacing device • Discontinuation of amiodaron or statin treatment during the follow-up period

Table 1: continued

Author	Population	Number of patients	Type of statin and dosage	Follow-up	Exclusion criteria
Kim, 2009	Permanent AF undergoing EC	81 (59.1 ± 10.5 years)	NS	13.1 ± 10.6 months	• Any previous EC • Significant mitral valve disease • Left atrium size > 55 mm • Recent infection • Surgery or acute coronary disease within 2 months of blood sample collection
Dogan, 2009	Persistent AF (< than 1 year) undergoing EC	221 (62.5 ± 8.9 years)	NS	ND	• Acute coronary syndrome • Severe valvular disease • Heart failure (NYHA > 2 class) • Left atrium size > 55 mm • Hepatic disfunction • Severe pulmonary disease • Hyperthyroidism • LV dysfunction (EF < 30 %) • RF
Baman, 2009	Persistent AF or atrial flutter undergoing EC	93	NS	15 months (10)	• Patients who have had > 1 ablation to eliminate to AF
Humphries, 2007	New onset AF undergoing EC	625 (mean age 63 years)	Atorvastatin, cerivastatin, sinvastatin, lovastatin, pravastatin	1 year	• Cardiothoracic surgery in previous 30 days • Missing information regarding hypertension history or medication use • Missed one year follow-up visit • Patients died • Patients withdrew from study before 1 year • Patients identified with chronic or permanent AF in the first visit after diagnosis
Watanabe, 2005	Symptomatic AF undergoing EC	106 (63 ± 14 years)	NS	140 ± 140 days	• Acute myocardial infarction • Unstable angina • Major surgical procedure within the previous month • Chronic obstructive pulmonary disease • Connective tissue disease • Acute infectious disease
Siu, 2003	Lone persistent AF (lasting > 3 months) undergoing EC	62 (61 ± 2 years)	4 sinvastatin (mean dose 20 ± 13 mg) 6 atorvastatin (mean dose 10 ± 3 mg) 32 ± 6 weeks before and 44 ± 1 months after	44 ± 1 m	• Structural heart disease • Hypertension • FA lasting > 3 months • Sepsis • Hyperthyroidism • Electrolyte inbalance
Koyama, 2009	Drug-refractory paroxysmal AF undergoing CA	186 59.7 (9.8)	NS	30 days	• Age >75 years, • Previous ablation • Persistent AF lasting >1 week • Hepatic or renal disease • Hyperthyroidism • Uncontrolled hypertension • LV dysfunction (ejection fraction < 45%) • Malignancy • Acute or chronic inflammatory disease
Park, 2009	Drug-refractory paroxysmal or persistent AF undergoing CA	152 NS	NS	18 months (14)	–
Richter, 2007	Patients with drug-resistant paroxysmal or persistent AF undergoing CA	234 56.7 (10.5)	Atorvastatin, sinvastatin, pravastatin, fluvastatin, rosuvastatin	12.7 months median (95%CI 11–14.4 m)	–
Al Chekakie, 2007	Patients with paroxysmal or persistent AF undergoing CA	177 NS	NS	13.8 months (8.6)	• Patients who underwent segmental ostial isolation or additional left atrial linear lesions

• AF, atrial fibrillation; CA, catheter ablation; CAD, coronary artery disease; EC, electrical cardioversion; EF, ejection fraction; LV, left ventricular; NS, not specified; ND, not declared.

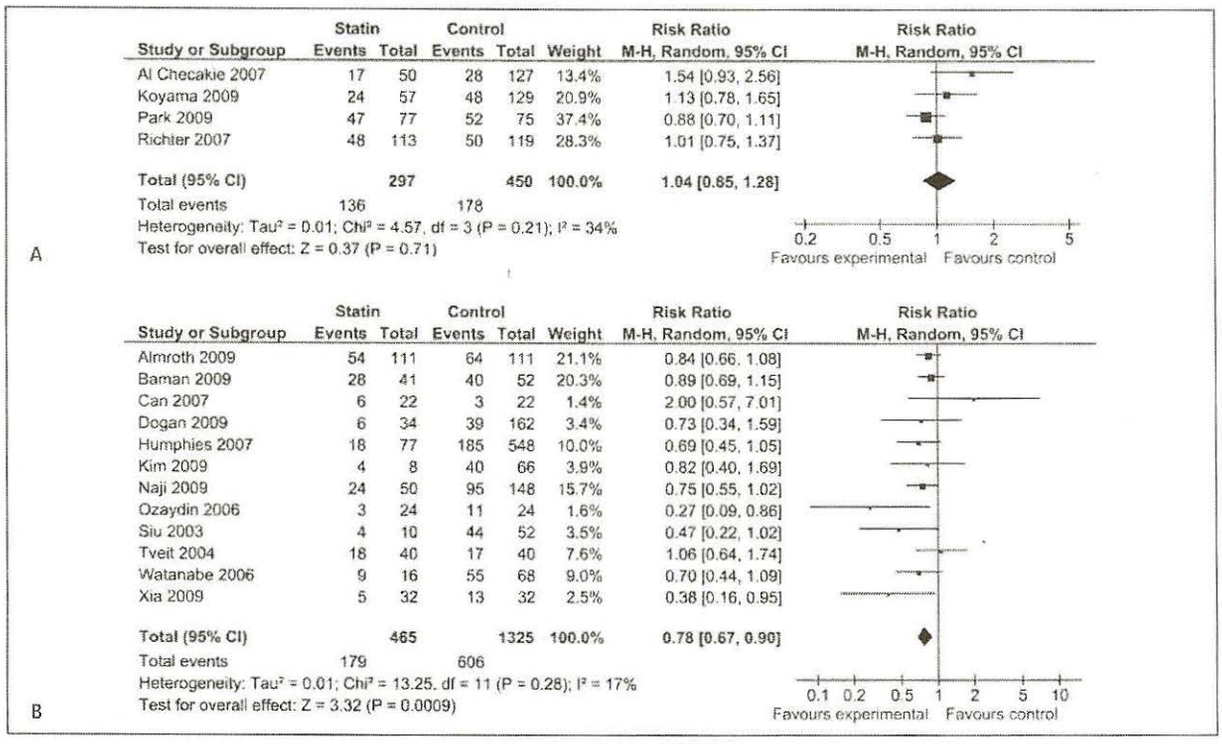


Figure 1: Pooled relative risk (RR) and 95% confidence intervals (CIs) of atrial fibrillation (AF) recurrence after catheter ablation (A) and electrical cardioversion (B).

Discussion

In our systematic review and meta-analysis of the literature we assessed the effect of statin therapy on the risk of AF recurrence after catheter ablation or electrical cardioversion. We found a statistically significant reduction in the incidence of recurrent AF after electrical cardioversion and no effect of statins on AF recurrence after catheter ablation. The 7.2% absolute risk reduction obtained with the use of statins after electrical cardioversion resulted in a number needed to treat of 14 patients. When the analysis was restricted to the five RCTs only, the magnitude of the effect was similar, but statistical significance was not reached. Although possibly due to the small number of patients included in these trials, lack of efficacy cannot definitively be excluded. In addition, the type and dose of statin drug also varied across studies, which raises additional questions of statin equipotency (45).

The results of our meta-analysis of studies conducted in patients undergoing electrical cardioversion are in keeping with the results of previous RCTs and meta-analyses which have shown that statins are effective in reducing the incidence of AF in patients undergoing cardiac surgery (19), after acute coronary syndromes (46), and in patients in sinus rhythm with a history of previous AF (46). There is a strong biological plausibility to support these findings. Rapidly growing evidence supports a relationship between AF and both cellular and plasma markers of inflammation including high-sensitivity CRP, interleukin-6, and interleukin-8. Inflammation may interfere with the structural and electrical properties of the atrial myocardium, creating a susceptible substrate for AF (12, 47). Furthermore, increased atrial oxidative stress may play an important role in inducing and maintaining AF (48, 49). Finally, recent data show that a decrease in endocardial NOS expression and atrial NO bioavailability directly contribute to the pathogenesis of AF (49), and that an imbalanced expression of iNOS/eNOS with nitric oxide overproduction could contribute to protein nitration and cardiomyocyte apoptosis in human AF (50). Statins have been shown to exhibit several vascular protective effects, including anti-inflammatory and antithrombotic properties, that are not related to changes in lipid profile (17, 51). Because statins can improve endothelial NO production, have anti-inflammatory effects and reduce oxidative stress, these drugs may act by preventing the establishment of a substrate for AF (51). Moreover, both Rac1 and RhoA, which are upregulated and mediate signal transduction integral to the pathogenesis of AF (52), are inhibited by statins.

We failed to observe a positive effect of statins in patients undergoing catheter ablation. The long duration of AF prior to ablation might account for the lack of efficacy of these medications since the patients are more likely to have established fibrosis and scarring, and thus less likely to respond to medications that inhibit inflammation. Furthermore, most of the clinical recurrences associated with this procedure are due to recovery of the lesion, mech-

What is known about this topic?

- Statins have important pleiotropic effects and have been shown to reduce vascular inflammation.
- Some evidence suggests that statins may have a role in the primary prevention of atrial fibrillation (AF), whereas little is know on the role of statins in patients with existing AF.

What does this paper add?

- The results of this study indicate that statins may lower the risk of AF recurrence after electrical cardioversion, but not after catheter ablation.
- Our study provides additional information to support the design of future randomised controlled trials to rigorously evaluate this question. Optimal dose selection and periods of follow-up appropriate to answer this question will be paramount.

anism that is independent by the action of these medications (53). On the other hand, this finding may be explained by the small number and by the relatively low quality of studies included, differences in study populations, and the retrospective study design in which temporal relationships between statin exposure and AF recurrence are difficult to discern. In addition, statin type and dose were not specified significantly limiting any informed conclusions. The effect of statins on AF recurrence following catheter ablation awaits results of well-designed RCTs.

Our meta-analysis has several limitations. First, our systematic review includes RCTs and observational studies. Application of formal meta-analytic methods to observational studies is controversial, since bias implicit in the study design may misrepresent the strength of associations within the data (22). To minimise this potential bias, we selected only studies in which the diagnosis of AF recurrence was objectively confirmed. Furthermore, we strictly followed the guidelines for quality of reporting of meta-analysis of RCTs and observational studies (PRISMA, MOOSE) to better clarify our results (21, 22). Second, the studies included in our meta-analysis had different inclusion and exclusion criteria. However, the heterogeneity among the studies, calculated using the I^2 statistic, was generally low. Third, since there were only a few studies assessing the role of statins in patients undergoing catheter ablation, the presence of publication bias in this setting could not be excluded. In patients undergoing electrical cardioversion, the funnel plot of RR versus standard error appeared slightly asymmetric with an absence of studies in the bottom right hand corner, suggesting that smaller, unpublished studies that demonstrate an increased RR of AF recurrence in patients taking statins may be not included in our meta-analysis. Our findings pertain to the effects of statins as a drug class. As stated previously, the type and dose of statin varied significantly across studies, e.g. atorvastatin dose ranged from 10 mg to 80 mg, raising questions of differential potency. Furthermore, due to the limitation of meta-analytic approach we were not able to explore possible additive effects of statins to other drugs with more established efficacy on secondary prevention of AF.

The prevalence of AF is projected to greatly increase over the next few decades particularly among older adults (54). The 30-day mortality related to AF stroke is 24% and haemorrhagic complications limit widespread use of anticoagulant therapy in the oldest and highest risk patients (55). Electrical cardioversion remains the most commonly used method to restore sinus rhythm, and despite a number of pharmacologic strategies to maintain sinus rhythm, a considerable proportion of patients continue to relapse to AF (1). The results of our meta-analysis suggest that statins may be effective in reducing AF recurrence after electrical cardioversion. Our study provides additional information to support the design of future RCTs to rigorously evaluate this question. Optimal dose selection and periods of follow-up appropriate to answer this question will be paramount.

Conflict of interest

None declared.

References

1. Krahn AD, Manfreda J, Tate RB, et al. The natural history of atrial fibrillation: incidence, risk factors, and prognosis in the Manitoba Follow-up Study. Am J Med 1995; 98: 476–484.
2. Goto S, Bhatt DL, Rother J, et al; REACH Registry Investigators. Prevalence, clinical profile, and cardiovascular outcomes of atrial fibrillation patients with atherothrombosis. Am Heart J 2008; 156: 855–863.
3. Marini C, Di Napoli M, Carolei A. Influence of cholesterol on survival after stroke. Effect of cholesterol on prognosis may rely on negative association with atrial fibrillation. Br Med J 1997; 315: 1159.
4. Chung MK. Randomized trials of rate vs rhythm control for atrial fibrillation. J Interv Card Electrophysiol 2004; 10: 43–53.
5. van Gelder IC, Hagens VE, Bosker HA, et al. A comparison of rate control and rhythm control in patients with recurrent persistent atrial fibrillation. N Engl J Med 2002; 347: 1834–1840.
6. Hohnloser SH, Kuck KH, Lilienthal J. Rhythm or rate control in atrial fibrillation Pharmacological Intervention in Atrial Fibrillation (PIAF): a randomised trial. Lancet 2000; 356: 1789–1794.
7. Wyse DG, Waldo AL, DiMarco JP, et al. A comparison of rate control and rhythm control in patients with atrial fibrillation. N Engl J Med 2002; 347: 1825–1833.
8. Opolski G, Torbicki A, Kosior DA, et al. Rate control vs. rhythm control in patients with nonvalvular persistent atrial fibrillation: the results of the Polish How to Treat Chronic Atrial Fibrillation (HOT CAFE) Study. Chest 2004; 126: 476–486.
9. Korantzopoulos P, Kolettis TM, Papathanasiou A, et al. Propafenone added to ibutilide increases conversion rates of persistent atrial fibrillation. Heart 2006; 92: 631–634.
10. Oral H, Pappone C, Chugh A, et al. Circumferential pulmonary vein ablation for chronic atrial fibrillation. N Engl J Med 2006; 354: 934–941.
11. Cappato R, Calkins H, Chen SA, et al. Worldwide survey on the methods, efficacy, and safety of catheter ablation for human atrial fibrillation. Circulation 2005; 111: 1100–1105.
12. Frustaci A, Chimenti C, Bellocci F, et al. Histological substrate of atrial biopsies in patients with lone atrial fibrillation. Circulation 1997; 96: 1180d.
13. Aviles RJ, Martin DO, Apperson-Hansen C, et al. Inflammation as a risk factor for atrial fibrillation. Circulation 2003; 108: 3006–3010.
14. Engelmann MDM, Svendsen JH. Inflammation in the genesis and perpetuation of atrial fibrillation. Eur Heart J 2005; 26: 2083–2092.
15. Liu T, Li G, Li L, Korantzopoulos P. Association between C-reactive protein and recurrence of atrial fibrillation after successful electrical cardioversion: a meta-analysis. J Am Coll Cardiol. 2007; 49: 1642–1648.
16. Jones SP, Teshima Y, Akao M, et al. Simvastatin attenuates oxidant-induced mitochondrial dysfunction in cardiac myocytes. Circ Res 2003; 93: 697–699.
17. Ieter DJ. Statins as potent antiinflammatory drugs. Circulation 2002; 106: 2041–2042.

18. Young-Xu Y, Jabbour S, Goldberg R, et al. Usefulness of statin drugs in protecting against atrial fibrillation in patients with coronary artery disease. Am J Cardiol 2003; 92: 1379–1383.
19. Liakopoulos OJ, Choi YH, Haldenwang PL, et al. Impact of preoperative statin therapy on adverse postoperative outcomes in patients undergoing cardiac surgery: a meta-analysis of over 30,000 patients. Eur Heart J 2008; 29: 1548–1559.
20. Fuster V, Ryden LE, Cannom DS, et al. ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation—executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Revise the 2001 Guidelines for the Management of Patients With Atrial Fibrillation). J Am Coll Cardiol 2006; 48: 854–906.
21. Moher D, Liberati A, Tetzlaff J, et al., PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Ann Intern Med 2009; 151: 264–269.
22. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. J Am Med Assoc 2000; 283: 2008–2012.
23. McGinn T, Wyer PC, Newman TB, et al.; Evidence-Based Medicine Teaching Tips Working Group.Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). CMAJ 2004; 171: 1369–1373.
24. Jadad AR, Moore AR, Carrol D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Controlled Clin Trials 1996; 17: 1–12.
25. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986; 7: 177–188.
26. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. Br Med J 2003; 327: 557–560.
27. Sterne JA, Egger M, Smith GD. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. Br Med J 2001; 323: 101–105.
28. García-Fernández A, Marín F, Mainar L, et al. Effect of statins on preventing recurrence of atrial fibrillation after electrical cardioversion. Am J Cardiol 2006; 98: 1299–1300.
29. Naji F, Suran D, Kanic V, et al. Statins and amiodarone improve freedom from recurrence of atrial fibrillation after successful cardioversion. Med Sci Monit. 2009; 15: CR494–498.
30. Almroth H, Hoglund N, Boman K, et al. Atorvastatin and persistent atrial fibrillation following cardioversion: a randomized placebo-controlled multicentre study. Eur Heart J 2009; 30: 827–833.
31. Xia W, Yin Z, Li J, et al. Effects of Rosuvastatin on Asymmetric Dimethylarginine Levels and Early Atrial Fibrillation Recurrence after Electrical Cardioversion. Pacing Clin Electrophysiol 2009; 32: 1562–1566.
32. Koyama T, Sekiguchi Y, Tada H, et al. Comparison of characteristics and significance of immediate versus early versus no recurrence of atrial fibrillation after catheter ablation. Am J Cardiol 2009; 103: 1249–1254.
33. Dogan A, Akcay S, Karabacak M, et al. The effect of pretreatment with renin-angiotensin-aldosterone system blockers on cardioversion success and acute recurrence of atrial fibrillation. Int J Clin Pract 2009; 63: 1017–1023.
34. Baman TS, Gupta SK, Billakanty SR, et al. Time to cardioversion of recurrent atrial arrhythmias after catheter ablation of atrial fibrillation and long-term clinical outcome. J Cardiovasc Electrophysiol 2009; 20: 1321–1325.
35. Park JH, Oh YS, Kim JH, et al. Effect of Angiotensin Converting Enzyme Inhibitors and Angiotensin Receptor Blockers on Patients Following Ablation of Atrial Fibrillation. Korean Circ J 2009; 39: 185–188.
36. Kim SK, Pak HN, Park JH, et al. Clinical and serological predictors for the recurrence of atrial fibrillation after electrical cardioversion. Europace 2009; 11: 1632–1638..
37. Humphries KH, Lee M, Sheldon R, et al; CARAF Investigators. Statin use and recurrence of atrial fibrillation after successful cardioversion. Am Heart J 2007; 154: 908–913.
38. Richter B, Derntl M, Marx M, et al. Therapy with angiotensin-converting enzyme inhibitors, angiotensin II receptor blockers, and statins: no effect on ablation outcome after ablation of atrial fibrillation. Am Heart J 2007; 153: 113–119.
39. Al Chekakie MO, Akar JG, Wang F, et al. The effects of statins and renin-angiotensin system blockers on atrial fibrillation recurrence following antral pulmonary vein isolation. J Cardiovasc Electrophysiol. 2007; 18: 942–946.
40. Ozaydin M, Varol E, Aslan SM, et al. Effect of atorvastatin on the recurrence rates of atrial fibrillation after electrical cardioversion. Am J Cardiol 2006; 97: 1490–1493.
41. Watanabe E, Arakawa T, Uchiyama T, et al. High-sensitivity C-reactive protein is predictive of successful cardioversion for atrial fibrillation and maintenance of sinus rhythm after conversion. Int J Cardiol 2006; 108: 346–353.
42. Tveit A, Grundtvig M, Gundersen T, et al. Analysis of pravastatin to prevent recurrence of atrial fibrillation after electrical cardioversion. Am J Cardiol 2004; 93: 780–782.
43. Siu CW, Lau CP, Tse HF. Prevention of atrial fibrillation recurrence by statin therapy in patients with lone atrial fibrillation after successful cardioversion. Am J Cardiol 2003; 92: 1343–1345.
44. Can I, Demir K, Ozdemir K, et al. Atorvastatin did not prevent recurrence of atrial fibrillation in addition to antiarrhythmic therapy after electrical cardioversion European Heart J 2007; 28 (Suppl.): Abstract 809.
45. Penning-van Beest F, Termorshuizen F, Goettsch W, et al. Adherence to evidence-based statin guidelines reduces the risk of hospitalizations for acute myocardial infarction by 40%: a cohort study. Eur Heart J 2007; 28: 151–159.
46. Fauchier L, Pierre B, de Labriolle A, et al. Antiarrhythmic effect of statin therapy and atrial fibrillation a meta-analysis of randomized controlled trials. J Am Coll Cardiol 2008; 51: 828–835.
47. Boos CJ, Anderson RA, Lip GY. Is atrial fibrillation an inflammatory disorder? Eur Heart J 2006; 27: 136–149.
48. Cai H, Li Z, Goette A, et al. Downregulation of endocardial nitric oxide synthase expression and nitric oxide production in atrial fibrillation: potential mechanisms for atrial thrombosis and stroke. Circulation 2002; 106: 2854–2858.
49. Mihm MJ, Yu F, Carnes CA, et al. Impaired myofibrillar energetics and oxidative injury during human atrial fibrillation. Circulation 2001; 104: 174–180.
50. Han W, Fu S, Wei N, et al. Nitric oxide overproduction derived from inducible nitric oxide synthase increases cardiomyocyte apoptosis in human atrial fibrillation. Int J Cardiol 2008; 130: 165–173..
51. Adam O, Neuberger HR, Böhm M, et al. Prevention of atrial fibrillation with 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors. Circulation 2008; 118: 1285–1293.
52. Adam O, Lavall D, Theobald K, et al. Rac1-induced connective tissue growth factor regulates connexin 43 and N-cadherin expression in atrial fibrillation. J Am Coll Cardiol 2010; 55: 469–480.
53. Wright M, Haissaguerre M, Knecht S, et al. State of the art: catheter ablation of atrial fibrillation. J Cardiovasc Electrophysiol. 2008; 19: 583–592.
54. Hylek EM, Evans-Molina C, Shea C, et al. Major Hemorrhage and Tolerability of Warfarin in the First Year of Therapy Among Elderly Patients with Atrial Fibrillation. Circulation 2007; 115: 2689–2696.
55. Hylek EM, Go AS, Chang Y, et al. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. New Engl J Med 2003; 349: 1019–1026.

Journal of Critical Care (2011) **26**, 431.e1–431.e9

ELSEVIER

Journal of Critical Care

Patient factors associated with transfusion practices in Veterans Affairs intensive care units: Implications for further research☆,☆☆,★

Yew Y. Ding MBBS, FRCP, MPH*, Boris Kader PhD, Cindy L. Christiansen PhD, Dan R. Berlowitz MD, MPH*

Center for Health Quality, Outcomes and Economic Research, Edith Nourse Rogers Memorial Hospital (152), Bedford, MA 01730, USA

Keywords:
Blood transfusion;
Intensive care units;
Anemia;
Risk adjustment;
Comorbidity;
Hospital mortality

Abstract

Purpose: We sought to describe how patient characteristics influence the frequency of red blood cell (RBC) transfusions among critically ill patients after taking into account hemoglobin (Hgb) level.

Methods: This was a retrospective cohort study using secondary analysis of administrative data of Veterans Affairs intensive care unit (ICU) admissions. The outcome of interest was RBC transfusion during the first 30 days of ICU admission. Besides Hgb level, explanatory variables included demographics, admission-related information, comorbid conditions, ICU admission diagnosis, and selected laboratory test results. Logistic regression modeling quantified associations between explanatory variables and transfusion.

Results: For 259 281 ICU admissions from 2001 to 2005, the overall incidence of RBC transfusion was 12.5%. Increased age, male gender, admission for acute myocardial infarction (AMI), and comorbid heart disease were independently associated with transfusion. Compared with admission for reference diagnoses, transfusions were more likely for admissions for AMI, unstable angina, and congestive heart failure only at Hgb levels below 11, 9, and 6 g/dL, respectively.

☆ Sponsors: the National Medical Research Council (NMRC) of Singapore, Ortho Biotech Clinical Affairs, and the Department of Veterans Affairs played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

☆☆ Competing interests: the authors declare that they have no competing interests.

★ Authors' contributions: Y.Y.D. contributed to the study design, data analyses and interpretation, and drafted the manuscript. B.K. acquired and analyzed the data, contributed to its interpretation, and revised the manuscript. C.C. contributed to the study design, data analysis and interpretation, and revised the manuscript. D.B. conceived the study, contributed to study design and data interpretation, and revised the manuscript. All authors read and approved the final manuscript.

* Corresponding authors. Yew Y. Ding is to be contacted for communications on the paper at Department of Geriatric Medicine, Tan Tock Seng Hospital, Singapore 308433. Dan R. Berlowitz is to be contacted for reprints at VA Bedford Hospital, Center for Health Quality, Outcomes and Economic Research, Edith Nourse Rogers Memorial Hospital (152), Bedford, MA 01730, and Boston University School of Public Health, Boston, MA 02118.

E-mail addresses: dyyttsh@gmail.com (Y.Y. Ding), dan.berlowitz@va.gov (D.R. Berlowitz).

0883-9441/$ – see front matter
doi:10.1016/j.jcrc.2010.12.012

Conclusions: Intensive care unit patients admitted for AMI, unstable angina, and congestive heart-failure had higher likelihood of receiving RBC transfusions below specific Hgb levels varying from 6 to 11 g/dL. Further research is needed to determine how these transfusion practices influence outcomes.

1. Introduction

Anemia is exceptionally common among patients admitted to an intensive care unit (ICU). Consequently, red blood cell (RBC) transfusions are extremely common, with some studies suggesting that this occurs among 20% to 53% of critically ill patients [1]. In a prospective, observational study involving 1136 patients from 145 western Europe ICUs, Vincent et al [2] found that 37% received RBC transfusions during the first 28 days of ICU admission. Older patients and those who stayed longer were more commonly transfused. In another prospective, observational cohort study of 4892 patients admitted to 284 ICUs in the United States, Corwin et al [3] found that 44% of patients had RBC transfusions during the first 30 days of ICU admission, with the number of transfusions received by a patient being independently associated with longer ICU stays. These studies had relatively small samples and examined practices in 2001 and earlier. Moreover, only a limited number of potential predictors of transfusion were studied. Although hemoglobin (Hgb) level is clearly an important determinant for transfusion, the decision to transfuse blood in ICU patients is often more complex. There is a need to better understand factors contributing to this decision beyond the Hgb level.

Although RBC transfusion has long been considered a relatively safe procedure, a growing body of literature suggests that it is associated with an increased incidence of nosocomial infections, acute lung injury, multiorgan failure, and mortality [4-8]. The benefits of Transfusions have also been questioned because of the Transfusion Requirements in Critical Care (TRICC) Trial, a landmark, randomized, controlled trial that demonstrated that critically ill patients managed with a liberal transfusion policy had similar outcomes with those managed with a restricted policy that advocated transfusions only at Hgb levels below 7 g/dL [9]. More than ever before, it is critical to identify factors beyond Hgb level that are associated with likelihood of transfusions to gain valuable insights on how physicians make their clinical decisions to transfuse in the ICU setting.

We therefore conducted an observational study to examine transfusion practices among ICU patients of Veterans Affairs (VA) medical centers. The large sample of patients available from the extensive electronic databases of the VA, which is the largest health delivery system in the United States, conferred a significant advantage in achieving our study objective, which was to identify patient characteristics associated with an increased likelihood of RBC transfusion after adjusting for Hgb level. We sought to test our hypothesis that specific comorbid conditions and admitting diagnoses are independently associated with higher likelihood of transfusion among ICU patients. Our results will identify important factors that influence the decision of physicians to transfuse and can then provide the basis for further research and quality improvement efforts.

2. Materials and methods

2.1. Study design and data sources

We conducted a retrospective cohort study using secondary analysis of VA electronic databases at the Center for Health Quality, Outcomes and Economic Research at VA Bedford in Massachusetts. Ethical approval was obtained from the VA Bedford Institutional Review Board, which also approved the study procedures. We obtained detailed clinical information at the national level through several VA databases. Information on each medical encounter was obtained from the National Patient Care Database. Selected laboratory results were extracted from the Decision Support System. The date of death was determined by combining information from the Beneficiary Identification Record Locator Subsystem, National Patient Care Database, and Social Security files.

2.2. Sample

We downloaded relevant data from the Veterans Health Administration Office of Information at the Austin Automation Center. The study database included information on all medical ICU admissions in the VA for the years 2001 through 2005. For hospitals that may have had separate surgical and neurologic ICUs, we only considered the medical ICU. We also excluded operative cases defined as having surgical admitting diagnoses. The purpose was to avoid accounting for postsurgical transfusion practice that probably has different determining factors. We only included the first ICU admission of each year for any given patient.

2.3. Dependent variable

The outcome (dependent) variable for this study was administration of blood transfusion during the first 30 days of ICU admission. A transfusion was considered as having been received if there was the *International Classification of*

Diseases, 9th Revision, Clinical Modification (*ICD-9-CM*) procedure code of 99.04 (transfusion of packed cells) or 99.03 (transfusion of whole blood) on a date in which the patient is known to be in the ICU and from the specific ICU file of the inpatient records. For patients with multiple dates on which transfusions were listed, each date was recorded. Intensive care unit admissions during which at least 1 transfusion had been documented during the first 30 days of stay were collectively assigned as the "transfusion" group. All other ICU admissions were assigned as the "no transfusion" group. To identify hospitals that may be systematically underreporting transfusions, we examined rates at each VA medical center. To exclude hospitals that were possible outliers with respect to transfusion practices, we arbitrarily removed data from 8 out of 120 hospitals where the transfusion rates were beyond 2 standard deviations from the mean rate.

2.4. Independent variables

Besides Hgb level, explanatory (independent) variables examined included (1) demographics: age, sex, race, or ethnicity; (2) admission-related information: year, hospital, source of admission to ICU (direct or transfer); (3) comorbid conditions: based on the set of Elixhauser et al [10] of 30 comorbidity measures, Acute Physiology and Chronic Health Evaluation (APACHE) III chronic health [11], and others [12,13]; (4) ICU admission diagnosis: based on medical diagnoses used by Health Cost and Utilization Project [14]; and (5) laboratory values of blood tests: serum creatinine and others based on APACHE III acute physiologic abnormalities [11].

Demographic variables such as sex and race were included based on clinical opinion. Admission-related variables such as year and medical center were added to take into account the potential clustering effects of time and hospital. Source of ICU admission was either "direct" (from sources outside the hospital) or "transfer" (from another unit in the hospital). This information may reflect the rapidity of illness onset, which in turn may influence likelihood of receiving transfusion.

Hemoglobin level has been shown to be strongly associated with performance of transfusion [2,3] and therefore would need to be adjusted for in the endeavor to isolate the effects of other patient factors. We developed an algorithm to select 1 Hgb value (designated as Hgb level) from among others for that ICU admission. First, we selected the lowest value during the period from ICU admission to blood transfusion within the first 30 days of ICU stay. If they were not transfused in that period, then we selected the lowest value from ICU admission to 30 days after ICU admission or to ICU discharge, whichever occurred earlier. If no value during the first 30 days of ICU admission was available, then we selected the lowest value during the period from day of hospital admission to day of ICU admission. The Hgb test closest to the transfusion was not selected because there was no corresponding test for admissions without transfusion. To account for any acute decrease in Hgb usually caused by hemorrhage, we created an additional variable called Hgb change, defined as the difference between the latest Hgb test before hospitalization and Hgb level.

Chronic comorbid diseases and ICU admission diagnoses may influence the likelihood of receiving transfusion and would also need to be adjusted for. The *ICD-9-CM* codes for comorbid conditions and ICU admission diagnoses were those used in the original study by Elixhauser et al [10] and adapted from those used by Health Cost and Utilization Project [14], respectively. Comorbid conditions were considered present if the *ICD-9-CM* code of the diagnosis was stated at least once in administrative records during the 2 years before the hospital admission. Where there was an overlapping admission diagnosis, the chronic comorbid disease was not assigned to avoid double counting of diagnostic information (Appendix A). To identify patients with chronic kidney disease, we used the following operational definition:

1. estimated glomerular filtration rate (eGFR) less than 60 mL/min per 1.73 m^2 on first day of ICU admission; and
2. eGFR less than 60 mL/min per 1.73 m^2 done at least 90 days before ICU admission.

These criteria were consistent with the Kidney Disease Outcomes Quality Initiative definition of chronic kidney disease that required patients to have glomerular filtration rate less than 60 mL/min per 1.73 m^2 for at least 3 months [15]. The eGFR values were calculated from the serum creatinine values using the predictive equation derived from the Modification of Diet in Renal Disease study [16]. The ICU admission diagnosis was obtained by selecting the first *ICD-9-CM* code from the ICU files.

Besides Hgb level, candidate blood tests based on APACHE III acute physiologic abnormalities [11] were white blood cell count, serum creatinine, blood urea nitrogen, serum sodium, serum albumin, serum bilirubin, and serum glucose. However, arterial blood gases were not included because they were not available from our data sources.

Other candidate independent variables that may be associated with administration of blood transfusion but were not available in the national VA data sets were functional status, cognitive scores, and clinical parameters.

Several studies have evaluated the quality of data included in VA national databases. In 3 studies, diagnoses listed in the database have compared favorably with medical records [17-19]. Data on race have also compared favorably to patient self-report [20].

2.5. Analyses

We excluded variables where there were missing values for more than 20% of the sample. For the remaining

variables, we deleted the entire record if there was a missing value. Although this was likely to result in decrease in the sample size available for the analysis and consequent loss of power, it was the simplest approach. Other commonly used methods, such as simple imputation, were not used because we could not assume that the data were missing completely at random. We compared baseline characteristics for included and excluded ICU admissions to alert us to any major differences between these 2 groups.

We performed principal components analyses to group together comorbidity variables that were highly correlated with each other (Appendix B). The purpose was to reduce the original 30 comorbid conditions to a smaller number of comorbidity groupings to facilitate parsimony in regression modeling. We performed logistic regression to build explanatory models for receipt of transfusion. In addition, we used hierarchical modeling to account for clustering within multiple years hospitals, and ICU admissions for unique patients. The purpose was to determine if treatment estimates and their confidence intervals differed significantly when patient clustering was taken into account. For comorbid heart disease and cardiac ICU admission diagnoses, we included interaction variables of these diagnoses with Hgb level in the model. This is to explore the possibility that Hgb level may be an effect modifier in the relationship between these conditions and transfusion receipt.

We used Statistical Analyses System (SAS Institute, Cary, NC) to perform the data analyses. Statistical significance was taken at $P < .05$.

3. Results

The total number of ICU admissions during the period of fiscal years from 2001 to 2005 was 302 059. After selecting only the first ICU admission for each fiscal year for any patient, the number of ICU admissions per unique patient was 1 (86.2%), 2 (11.4%), 3 (2.0%), 4 (0.4%), and 5 (0.1%). The mean age (SD) of patients was 66.0 years (12.5). Only 2.8% of ICU admissions were for female patients, as was expected for the VA. Ethnic distribution of patients was white (62.7%), black (17.1%), Hispanic (4.5%), and others (15.7%). Rather than transfer from another inpatient hospital unit, direct admission to the ICU from sources outside the hospital occurred in 76.8%. The mean baseline Hgb level (SD) was 11.4 g/dL (2.6).

Due to missing values, 42,778 (14.2%) ICU admissions were excluded, leaving 259,281 available for multivariable analyses. Hgb level and almost all other variables were similar for both included and excluded admissions (Appendix C). The incidence of RBC transfusion during the first 30 days of ICU admission was 12.5%. The mean time (SD) from ICU admission to first blood transfusion received was 1.8 days (3.4).

Patient characteristics of transfused and nontransfused patients are compared in Table 1. Principal components

Table 1 Comparison of the characteristics of transfused with nontransfused patients (N = 259 281)

Explanatory variables	Transfused (n = 32 386)	Nontransfused (n = 226 895)
Hgb level (g/dL)	8.1	11.9
Demographic and admission features:		
Age (y)	68.1	65.8
Male (%)	97.7	97.1
White race (%)	60.6	62.8
Black race (%)	18.0	16.6
Other races (%)	7.0	5.4
Unknown race (%)	14.4	15.2
Direct admission (%)	69.5	76.9
ICU length of stay (d)	7.2	4.2
Chronic kidney disease (%)	29.1	20.5
Comorbidity groupings		
Heart disease (%)	42.5	34.6
Heart disease risk factors (%)	66.8	58.5
Neurologic disease (%)	10.4	9.7
Respiratory disease (%)	30.7	30.5
Liver disease (%)	10.4	5.8
Psychiatric disease (%)	35.5	35.5
Coagulopathy (%)	10.5	5.2
Renal disease (%)	18.4	11.6
Anemia (%)	42.4	20.3
Cancer (%)	25.0	17.9
Peptic ulcer disease (%)	6.0	4.4
Endocrine disease (%)	7.1	6.7
Rheumatological disease (%)	3.6	2.9
ICU admission diagnoses		
Angina/unstable angina (%)	3.2	0.6
CHF (%)	3.1	5.8
Arrhythmia (%)	1.1	4.25
AMI (%)	7.8	11.1
Other cardiovascular diseases (%)	5.7	16.0
Peripheral vascular disease (%)	0.4	0.4
Valvular heart disease (%)	0.3	0.4
Gastrointestinal bleeding (%)	26.6	2.4
Other GI disease (%)	5.2	2.8
Gastrointestinal neoplasm (%)	1.4	0.5
Serious neurologic disease (%)	1.4	3.6
Minor neurologic disease (%)	0.1	0.3
Other medical diagnoses (%)	24.3	28.7
Orthopedic, nonsurgically treated (%)	1.3	0.9
Renal disease (%)	3.4	2.2
Lung neoplasm (%)	0.8	0.8
Respiratory arrest/failure (%)	5.3	3.6
Pneumonia (%)	4.1	4.1
Chronic obstructive pulmonary disease (%)	1.0	3.7
Other respiratory disease (%)	1.2	1.8
Infection (%)	5.1	3.7
Laboratory test results		
Serum creatinine (mg/dL)	2.2	1.7

Note: The 13 comorbidity groupings were obtained from the original 30 comorbid conditions through principal components analyses. Differences between transfused and nontransfused patients for all explanatory variables were statistically significant at $P < .05$ except for the comorbidity groupings, respiratory disease, and psychiatric disease.

Table 2 Regression of RBC transfusion on baseline Hgb level and control variables

Explanatory variables	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Hgb level (g/dL)	0.44 (0.44-0.45) *	0.47 (0.47-0.47) *
Age (y)	1.02 (1.01-1.02) *	1.01 (1.00-1.01) *
Male (reference: female)	1.23 (1.14-1.33) *	1.19 (1.08-1.31) *
Race (reference: white)		
Black	1.12 (1.07-1.13) *	0.68 (0.65-0.70) *
Others	1.34 (1.28-1.40) *	1.12 (1.05-1.19) *
Unknown	0.99 (0.96-1.02)	0.95 (0.91-0.99) *
Direct admission (reference: transfer)	0.69 (0.67-0.70) *	1.07 (1.04-1.11) *
ICU length of stay (d)	1.06 (1.06-1.06) *	1.02 (1.02-1.02) *
Fiscal year (reference: 2001)		
2002	1.04 (1.00-1.08) *	1.00 (0.95-1.05)
2003	1.06 (1.02-1.10) *	1.00 (0.95-1.04)
2004	1.08 (1.04-1.12) *	0.98 (0.93-1.03)
2005	1.08 (1.04-1.12) *	0.99 (0.94-1.04)
Chronic kidney disease	1.60 (1.56-1.64) *	1.05 (1.01-1.09) *
Comorbidity groupings		
Cardiac disease	1.39 (1.36-1.43) *	1.06 (1.02-1.09) *
Cardiac risk factors	1.43 (1.39-1.46) *	0.99 (0.96-1.03)
Neurologic disease	1.07 (1.03-1.11) *	0.87 (0.82-0.91) *
Respiratory disease	1.01 (0.98-1.03)	0.96 (0.92-0.99) *
Liver disease	1.88 (1.81-1.96) *	1.07 (1.01-1.13) *
Psychiatric disease	1.00 (0.98-1.02)	0.96 (0.93-0.99) *
Coagulopathy	2.17 (2.08-2.26) *	1.18 (1.12-1.24) *
Renal disease	1.72 (1.67-1.77) *	0.96 (0.93-0.99)
Anemia	2.89 (2.82-2.96) *	1.14 (1.10-1.18) *
Peptic ulcer disease	1.40 (1.33-1.47) *	1.11 (1.04-1.18) *
Cancer	1.53 (1.49-1.58) *	1.03 (1.00-1.07)
Endocrine disease	1.07 (1.02-1.12) *	0.96 (0.90-1.01)
Rheumatologic disease	1.22 (1.15-1.30) *	1.01 (0.94-1.10)
ICU admission diagnoses (reference: other medical diagnoses)		
Angina/Unstable angina	0.24 (0.21-0.27) *	0.77 (0.66-0.90) *
CHF	0.63 (0.59-0.68) *	0.72 (0.66-0.78) *
Arrhythmia	0.30 (0.27-0.34) *	0.59 (0.52-0.66) *
AMI	0.83 (0.79-0.87) *	1.28 (1.21-1.36) *
Other cardiovascular diseases	0.42 (0.40-0.44) *	0.85 (0.80-0.90) *
Peripheral vascular disease	1.30 (1.09-1.56) *	1.00 (0.81-1.24)
Valvular heart disease	0.75 (0.59-0.94) *	0.63 (0.48-0.81) *
Gastrointestinal bleeding	13.01 (12.49-13.56) *	5.44 (5.17-5.73) *
Other GI disease	2.16 (2.04-2.29) *	1.55 (1.45-1.67) *
Gastrointestinal neoplasm	3.40 (3.04-3.80) *	1.57 (1.37-1.80) *
Serious neurologic disease	0.45 (0.40-0.49) *	0.55 (0.50-0.62) *
Minor neurologic disease	0.21 (0.13-0.35) *	0.37 (0.21-0.65) *
Orthopedic, nonsurgically treated	1.69 (1.52-1.89) *	1.22 (1.08-1.39) *
Renal disease	1.82 (1.70-1.95) *	0.86 (0.78-0.93) *
Lung neoplasm	1.22 (1.07-1.40) *	0.79 (0.68-0.93) *
Respiratory arrest/failure	1.72 (1.63-1.83) *	0.83 (0.78-0.89) *
Pneumonia	1.18 (1.11-1.25) *	0.77 (0.72-0.83) *
Chronic obstructive pulmonary disease	0.32 (0.29-0.36) *	0.48 (0.42-0.54) *
Other respiratory disease	0.82 (0.74-0.91) *	0.80 (0.70-0.90) *
Infection	1.63 (1.54-1.73) *	0.79 (0.74-0.85) *
Laboratory test results:		
Serum creatinine (mg/dL)	1.13 (1.12-1.14) *	1.01 (1.00-1.02)
Summary statistics		
c-Statistic	–	0.92
n	259 281	259 281

Dependent variable: receipt of RBC transfusion during the first 30 days of ICU admission.

Notes: The 13 comorbidity groupings were obtained from the original 30 comorbid conditions through principal components analyses. CI indicates confidence interval.

* Statistical significance at $P < .05$.

Table 3 Interaction effects in regression of RBC transfusion on baseline Hgb level and control variables

Explanatory variables	β coefficient estimates	SE	P
AMI	2.44	0.17	<.01
AMI * Hgb level	−0.24	0.02	<.01
Angina/unstable angina	4.09	0.55	<.01
Angina/unstable angina * Hgb level	−0.46	0.06	<.01
CHF	2.33	−0.30	<.01
CHF * Hgb level	−0.30	0.03	<.01
Comorbid heart disease	0.10	0.09	.24
Comorbid heart disease * Hgb level	−0.01	0.01	.49

Dependent variable: receipt of RBC transfusion during the first 30 days of ICU admission.

* Interaction between the 2 variables.

analyses permitted us to reduce the original 30 comorbid condition variables to the 13 grouping variables listed. Laboratory test variables, besides serum creatinine and Hgb level, were dropped on account of missing values for more than 20% of the sample population. Statistically significant differences were found for most of these variables. The unadjusted odds ratios (ORs) are presented in the second column of Table 2. Similarly, the unadjusted OR for ICU admission diagnoses where the reference group was admission diagnoses other than the 21 categories specified are also provided in that column.

After adjustment for Hgb level using logistic regression, there were several patient characteristics identified to be independently associated with the administration of RBC transfusion (Table 2). Intensive care unit admission for acute myocardial infarction (AMI) and comorbid heart disease increased the likelihood of transfusion. Other characteristics independently associated with transfusion were increasing age, races other than white or black, direct transfer to the ICU, increasing length of ICU stay, and ICU admission diagnoses of gastrointestinal (GI) bleeding, GI neoplasm, other GI disorders, and orthopedic problems. Additional comorbid conditions independently associated with transfusion included heart disease, liver disease, coagulopathy, anemia, peptic ulcer disease, and chronic kidney disease.

When Hgb level was substituted with Hgb change in the multivariable model, the c-statistic decreased from 0.92 to 0.87. When Hgb change was included as an additional variable, it was not independently associated with transfusion. In view of these findings, Hgb change was not selected for our final model.

The interactions of ICU admission diagnosis of AMI, unstable angina, and congestive heart failure (CHF) with Hgb level were all statistically significant, as shown in Table 3. The negative sign in the coefficient estimate for the interaction variable indicates that when any of these 3 conditions was the reason for ICU admission, the relative increase in likelihood of transfusion was more pronounced

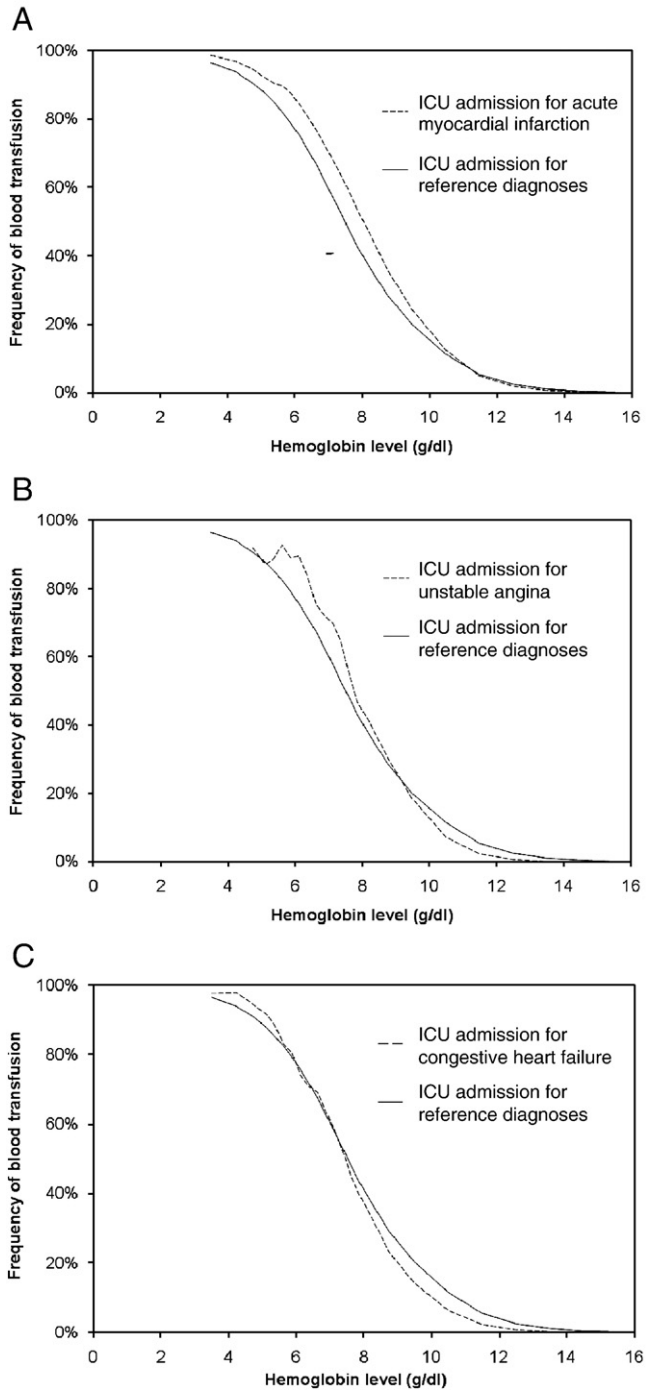


Fig. 1 Relationship between RBC transfusion and Hgb level using model predicted values for ICU admission diagnosis of AMI (A), unstable angina (B), and CHF (C) compared with ICU admission for reference diagnoses.

with decreasing Hgb levels. The interaction of comorbid heart disease and Hgb level was not statistically significant. The graph showing the model-predicted probability of receiving transfusions in Fig. 1A illustrates that ICU admissions for AMI had an increase in likelihood of transfusion when Hgb level was below about 11 g/dL, relative to those for reference medical diagnoses. Similarly,

information for ICU admissions for unstable angina is shown in Fig. 1B, indicating that relative increase in transfusions occurred when Hgb level was below about 9 g/dL. Finally, in the case of ICU admissions for CHF, relative increase in transfusions occurred only when Hgb level was below 6 g/dL. Additional details of this last subgroup are found in Appendix D.

Hierarchical modeling that accounted for clustering of ICU admissions within multiple years of data, multiple hospitals, and more than one ICU admission for individual patients did not result in significant changes in coefficient estimates (results not shown). The intra-class correlation coefficient for the random effects model accounting for clustering around hospitals was 0.25. This indicates that the proportion of transfusion variability explained by hospitals was 25%.

4. Discussion

To a large extent, Hgb level thresholds drive transfusion practices in critically ill patients [21]. In addition, a variable degree of influence is also exerted by other patient factors. In a scenario-based national survey of Canadian critical care practitioners, Hebert et al [22] and the TRICC investigators found that 75% of respondents selected an overall transfusion threshold of Hgb level 9 or 10 g/dL. Where there was GI hemorrhage or AMI, Hgb level 10 g/dL was the most frequently chosen threshold. Increased age, APACHE II score, hypoxemia, shock, coronary ischemia, and chronic anemia also raised transfusion thresholds, whereas CHF did not. Moving on from survey responses to actual clinical practice, the same investigators examined transfusion practice in a multicenter cohort study of 5298 tertiary-level ICU patients. They found that independent predictors of transfusion thresholds were age, APACHE II score, and institution [23]. Subsequently, in a prospective observational study of 3534 patients from 146 western European ICUs, Vincent observed that transfused patients were older and had higher admitting APACHE II scores and lower admitting Hgb levels. The overall mean pretransfusion Hgb level was 8.4 g/dL, with presence of coronary artery disease raising this level marginally to 8.7 g/dL [2]. In the CRIT (Anemia and blood transfusion in the critically ill – Current clinical practice in the United States) cohort study of 4892 patients from 284 American ICUs who were followed prospectively, Corwin et al [3] found that the mean pretransfusion Hgb level was 8.7 g/dL. There was little association of age with transfusion, and the incidence of transfusion was relatively consistent across comorbidities with the exception of anemia and hematologic disease being associated with more frequent transfusions. There were also more transfusions with the admitting diagnosis of GI hemorrhage and with higher APACHE II scores [3]. A recent study of 238 patients admitted to 5 general Israeli ICUs found that those with AMI had significantly higher Hgb level transfusion trigger (8.8 g/dL) than that for the whole group (7.9 g/dL). This trend was also seen for those with history of heart disease [24]. Although these studies were informative on the influence of specific patient characteristics, there is a need to examine how these factors operate while taking into account Hgb level and other patient factors. We argue that adopting this approach will provide insights on the way physicians think about transfusions in clinical practice. To this end, multivariable regression is needed to adjust for Hgb level and other patient factors so as to isolate the effect of any individual factor. Given the wide range of candidate patient factors to work on in regression analyses, a large sample size is required.

To the best of our knowledge, this is by far the largest observational study examining transfusion practices in ICUs. Using a large sample built on data from numerous ICUs across the United States albeit exclusively from the VA health care system, we found that age, male gender, admission for AMI, comorbid cardiac disease, and chronic kidney disease were independently associated with transfusion after adjustment for Hgb level and other measured patient factors. Not surprisingly, this was also the case for comorbid liver disease, coagulopathy, peptic ulcer disease, and anemia, as well as ICU admission for GI and orthopedic diagnoses. It is noteworthy that with the exception of admission for GI bleeding (OR, 5.44), these factors had relatively modest strengths of association (OR between 1.05 and 1.57). This reflects the complex relationships between multiple factors and their association with transfusion. However, demonstration that Hgb level was a significant effect modifier in the relationships between 3 cardiac ICU admission diagnoses and transfusion receipt is perhaps the most important contribution that our study brings to the expanding literature on transfusion practices in the critical care setting.

Relative to reference diagnoses, ICU admission for AMI was associated with higher likelihood of receiving transfusions at Hgb levels below about 11 g/dL. It is conventional clinical wisdom that transfusions will benefit patients with acute heart disease more as Hgb level decreases through facilitation of optimal oxygen delivery. It is quite plausible that this expectation is reflected in clinical practice. This strategy is supported by evidence from a large observational study on 78 974 older adults with AMI in which Wu et al [25] found that transfusions were significantly associated with lower mortality for hematocrit levels up to 33% (or Hgb level of 11 g/dL). With respect to 30-day mortality, the adjusted OR for transfusion was 0.22 for those with hematocrit below 24% (or Hgb level of 8 g/dL or lower) and 0.69 for those with hematocrit of 30.1% to 33.0% (Hgb level of 10.1 to 11.0 g/dL) [25]. Interestingly, a subsequent subgroup analysis of the TRICC trial focusing on 257 patients with ischemic heart disease found that those assigned to the "liberal" strategy of transfusing when Hgb level was below 10 g/dL had a nonstatistically significant reduction in 30-day mortality (21% vs 26%) compared with those who were assigned the "restrictive" strategy of transfusing only when

the Hgb level was below 7 g/dL [26]. Aronson et al [27] demonstrated that transfusion conferred a statistically significant reduction in mortality in another observational study on adults with Hgb levels of 8 g/dL or lower with adjusted hazard ratio of 0.13. Similarly, Alexander et al [28] found that for patients with non–ST elevation acute coronary syndromes including AMI experienced an almost significant trend toward inhospital mortality reduction (OR, 0.67) with transfusion when hematocrit was 24% or lower (or Hgb level of ≤8 g/dL). In contrast, analysis of data from 3 large international trials on acute coronary syndromes by Rao et al [29] showed that transfusion was significantly associated with increased mortality across all categories of hematocrit levels with adjusted hazard ratio of 3.54. Given the somewhat conflicting evidence on the effect of transfusion on mortality in the setting of AMI, further clinical trials are needed to determine the Hgb level below which transfusion will benefit patients rather than harm them. It is telling that although the restrictive transfusion strategy was not recommended for patients with AMI and unstable angina in a recently published clinical practice guideline on RBC transfusions in critical care, no specific guidance on an appropriate Hgb level threshold was offered [30]. Ideally, a randomized, controlled trial like the TRICC focusing specifically on patients with AMI in ICU will be needed. In the absence of such trials, a large and robust observational trial may shed more light on this issue. Given the knowledge of transfusion practices gained from our study, it is important and arguably urgent to obtain the answer to this question.

In contrast, ICU admission for unstable angina was associated with lower likelihood of transfusion compared with admission for reference admission diagnoses. However, transfusions became more likely at Hgb levels below about 9 g/dL. This level is slightly higher than the level (equivalent to Hgb level ≤8 g/dL) below which transfusions had a beneficial impact on mortality among those with non–ST elevation acute coronary syndromes including unstable angina in the study by Alexander [28]. Given the paucity of evidence on the Hgb level threshold for transfusions in the setting of unstable angina, additional clinical trials are needed to provide the answer.

Similarly, ICU admission for CHF was associated with lower likelihood of transfusion compared with reference admission diagnoses. This is not surprising, given that physicians are usually conservative on transfusions in this setting because of concern that transfusions may exacerbate the fluid overload state, preferring to transfuse only when Hgb levels are much lower. Indeed, our study showed that there was higher likelihood of transfusion relative to reference admission diagnoses only at Hgb levels of about 6 g/dL or lower. We are not aware of any published clinical trials that specifically address transfusion thresholds in the setting of CHF.

Compared with the studies by Vincent et al [2] and Corwin et al [3] where 37% and 44% of patients were transfused during the first month of ICU admission, only 12% of our study population received transfusions. It is likely that selection issues have, in part, accounted for these differences. Transfusions tend to be more common among surgical patients. Although the 2 earlier studies included surgical ICUs, we attempted to focus on medical ICUs. We excluded admissions to surgical and neurologic ICUs where these facilities were separate from the medical ICU at that hospital. We also excluded those admissions for diagnoses that were surgical in nature. It seems unlikely that transfusion threshold differences may have contributed to lower transfusion rates in our study, given that the mean pretransfusion Hgb level across all 3 studies were similar and within the 8 to 9 g/dL range.

The ICUs in our study were heterogeneous with respect to whether they were staffed only by attending physicians or had additional residents and fellows, teaching hospitals or not, and located in urban or rural areas. There are no national VA policies on transfusion decisions. Difference in hospitals probably had small to moderate influence on transfusion practice. This is reflected by the proportion of transfusion variability explained by hospitals being 25%. A recent study exploring hospital variation in transfusion among patients after cardiac surgery obtained similar findings [31].

There are 3 important limitations to our study. Firstly, no VA data exist on the accuracy of *ICD-9-CM* codes indicating transfusion receipt. However, among patients with lower Hgb levels, we found that up to about 70% received transfusions (results not shown). This suggests that most transfusions would probably have been identified using these codes. We also relied on external evidence from a previous study on validity of the billing for blood transfusion at one non-VA tertiary care hospital in the United States. Using the relevant ICD-9-CM procedure code, Segal found that sensitivity and specificity for identification of transfusions from blood bank records were 83 percent and 100 percent respectively [32]. Errors in transfusion timing are minimized because we examined the date and time of transfusion from procedure codes, and only identified transfusion codes found in the ICU files. While we acknowledge the possibility of under-coding of transfusions, we argue that its extent was likely to be relatively small. Transfusion codes were also used in a study of VA patients with chronic kidney disease [33]. Secondly, only about 3% of patients were female. However, given that the number of ICU admissions of female patients is in excess of 7000 and that we adjusted for sex in our multivariable regression analyses, we argue that our model is likely to be valid for both sexes. Finally, we did not have information on all dimensions of acute illness severity nor on functional status. We are uncertain on the extent to which the absence of this information may have introduced hidden bias. Given the high level of model discrimination (c-statistic >0.9) achieved, we believe that any unmeasured factors are unlikely to have significant impact on transfusion practice.

5. Conclusions

Independent of Hgb level and other patient factors, ICU admission diagnoses of AMI and comorbid heart disease increased likelihood of RBC transfusion. Compared with patients admitted for reference admitting diagnoses, those admitted for AMI, unstable angina, and CHF were more likely to be transfused only at Hgb levels below about 11, 9, and 6 g/dL, respectively. Given our findings, further research is needed to determine how these transfusion practices influence outcomes of patients admitted to the ICU.

Acknowledgment

Y.Y.D. was fully funded for his research fellowship at Boston by the National Medical Research Council of Singapore. This research project was supported, in part, by a grant from Ortho Biotech Clinical Affairs. The Department of VA provided space and computer support for the study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jcrc.2010.12.012.

References

[1] Vincent JL, Piagnerelli M. Transfusion in the intensive care unit. Crit Care 2006;34(Suppl):S96-101.
[2] Vincent JL, Baron J, Reinhart K, et al. Anemia and blood transfusion in critically ill patients. JAMA 2002;288:1499-507.
[3] Corwin HL, Gettinger A, Pearl RG, et al. The CRIT study: anemia and blood transfusion in the critically ill—current clinical practice in the United States. Crit Care Med 2004;32:39-52.
[4] Claridge JA, Sawyer RG, Schulman AM, et al. Blood transfusions correlate with infections in trauma patients in a dose-dependent manner. Am Surg 2002;68:566-72.
[5] Hill GE, Frawley WH, Griffith KE, et al. Allogenic blood transfusions increase the risk of postoperative bacterial infection: a meta-analysis. J Trauma 2003;54:908-14.
[6] Jawa RS, Anillo S, Kulaylat MN. Transfusion-related acute lung injury. J Intensive Care Med 2008;23:109-21.
[7] Moore FA, Moore EE, Sauaia M. Blood transfusion: an independent risk factor for postinjury multiple organ failure. Arch Surg 1997;132: 620-5.
[8] Taylor RW, Manganaro L, O'Brien J, et al. Impact of allogenic packed red blood cell transfusion on nosocomial infection rates in the critically ill patient. Crit Care Med 2002;30:2249-54.
[9] Hebert PC, Wells G, Blajachman MA, et al, and the Blood Transfusion Requirements in Critical Care Investigators for the Canadian Critical Care Trials Group. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. N Engl J Med 1999;340: 409-17.
[10] Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. Med Care 1998;36:8-27.
[11] Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults. Chest 1991;100:1619-36.
[12] Johnston JA, Wagner DP, Timmons S, et al. Impact of different measures of comorbid disease on predicted mortality of intensive care unit patients. Med Care 2002;40:929-40.
[13] Render ML, Kim HM, Welsh D, et al, for the VA ICU Project (VIP) Investigators. Automated intensive care unit risk adjustment: results from a National Veterans Affairs study. Crit Care Med 2003;31:1638-46.
[14] Healthcare Cost and Utilization Project (HCUP): Clinical Classifications Software (CCS) for *ICD-9-CM*, 2006.
[15] NKF: KDOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Am J Kidney Dis 2002;39 (2 suppl 1):S1-266.
[16] Levey AS, Bosch JP, Lewis JB, et al. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Int Med 1999;130:461-70.
[17] Borzecki AM, Wong AT, Hickey EC, et al. Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? Am J Med Qual 2004;19:201-6.
[18] Kashner TM. Agreement between administrative files and written medical records: a case of the department of Veterans Affairs. Med Care 1998;36:1324-36.
[19] Szeto HC, Coleman RK, Gholami P, et al. Accuracy of computerized outpatient diagnoses in a Veterans Affairs general medicine clinic. Am J Manag Care 2002;8:37-43.
[20] Boehmer U, Kressin NR, Berlowitz DR, et al. Self-Reported vs administrative race/ethnicity data and study results. Am J Public Health 2002;92:1471-3.
[21] Reiles E, Van der Linden P. Transfusion trigger in critically ill patients: has the puzzle been completed? Crit Care 2007;11:142-4.
[22] Hebert PC, Wells G, Martin C, et al, for the Transfusion Requirements in Critical Care Investigators and the Canadian Critical Care Trials Group. A Canadian survey of transfusion practices in critically ill patients. Crit Care Med 1998;26:482-7.
[23] Hebert PC, Wells G, Martin C, et al, for the Transfusion Requirements in Critical Care (TRICC) Investigators and the Canadian Critical Care Trials Group. Variation in red cell transfusion practice in the intensive care: a multicentre cohort study. Crit Care 1999;3:57-63.
[24] Cohen J, Kagan I, Hershcovici R, et al. Red cell transfusions—are we narrowing the evidence-practice gap? An observational study in 5 Israeli intensive care units. J Crit Care, doi:10.1016/j.jcrc.2010.03.010.
[25] Wu WC, Rathore SS, Wang Y, et al. Blood transfusion in elderly patients with acute myocardial infarction. N Engl J Med 2001;345:1230-6.
[26] Hebert PC, Yetisir E, Martin C, et al. Is a low transfusion threshold safe in critically ill patients with cardiovascular diseases? Crit Care Med 2001;29:227-34.
[27] Aronson D, Dann EJ, Bonstein L, et al. Impact of red blood cell transfusion on clinical outcomes in patients with acute myocardial infarction. Am J Cardiol 2008;102:115-9.
[28] Alexander MD, Chen AY, Wang TY, et al, for the CRUSADE Investigators. Transfusion practice and outcomes in non–ST-segment elevation acute coronary syndromes. Am Heart J 2008;155:1047-53.
[29] Rao SV, Jollis JG, Harrington RA, et al. Relationship of blood transfusion and clinical outcomes in patients with acute coronary syndromes. JAMA 2004;292:1555-62.
[30] Napolitano LM, Kurek S, Luchette FA, for the American College of Critical Care Medicine of the Society of Critical Care Medicine and the Eastern Association for the Surgery of Trauma Practice Management Workgroup. Clinical practice guideline: red blood cell transfusion in adult trauma and critical care. Crit Care Med 2009;37:3124-57.
[31] Rogers MAM, Blumberg N, Saint S, et al. Hospital variation in transfusion and infection after cardiac surgery: a cohort study. BMC Medicine 2009;7:37.
[32] Segal JB, Ness PM, Powe NR. Validating billing data for RBC transfusion: a brief report. Transfusion 2001;41:530-3.
[33] Lawler EV, Bradbury BD, Fonda JR, et al. Transfusion Burden among Patients with Chronic Kidney Disease and Anemia. Clin J Am Soc Nephrol 2010;5:667-72.

American Heart Association Atrial Fibrillation Research Summit

A Conference Report From the American Heart Association

N.A. Mark Estes III, MD, FAHA, Chair; Ralph L. Sacco, MS, MD, FAHA; Sana M. Al-Khatib, MD, MHS; Patrick T. Ellinor, MD, PhD, FAHA; Judy Bezanson, DSN, CNS, RN; Alvaro Alonso, MD, MPH; Charles Antzelevitch, PhD, FAHA; Randall G. Brockman, MD; Peng-Sheng Chen, MD; Sumeet S. Chugh, MD, FAHA; Anne B. Curtis, MD, FAHA; John P. DiMarco, MD, PhD; Kenneth A. Ellenbogen, MD, FAHA; Andrew E. Epstein, MD, FAHA; Michael D. Ezekowitz, MBChB, DPhil, FAHA; Pierre Fayad, MD, FAHA; Brian F. Gage, MD; Alan S. Go, MD; Mark A. Hlatky, MD, FAHA; Elaine M. Hylek, MD, MPH; Michael Jerosch-Herold, PhD; Marvin A. Konstam, MD; Richard Lee, MD; Douglas L. Packer, MD, FAHA; Sunny S. Po, MD, PhD; Eric N. Prystowsky, MD, FAHA; Susan Redline, MD, MPH; Yves Rosenberg, MD, MPH; David R. Van Wagoner, PhD, FAHA; Kathryn A. Wood, PhD, RN; Lixia Yue, PhD; Emelia J. Benjamin, MD, ScM, FAHA

Atrial fibrillation (AF) poses a major global public health challenge because it is increasing in prevalence and is associated with an increased risk of stroke, dementia, heart failure, and death.[1–3] In response to the many challenges posed by AF, the American Heart Association (AHA) convened a conference in Washington, DC, on June 12–13, 2010, that included patients, nurses, physicians, and healthcare policy makers and regulators. In addition, basic, translational, population, outcomes, and clinical scientists participated (Appendix). The 22 presentations and 6 panel discussions were organized into 4 sessions: (1) Mechanisms of AF: Basic and Translational Science and Genetics; (2) Epidemiology, Outcomes, Cost, AF, and Stroke Prevention; (3) Meeting the Clinical Challenges in AF; and (4) Redefining the Therapeutic Goals of AF (Appendix). The focus of the present report is to provide an overview of the key concepts presented and the core recommendations made by the summit participants.

Mechanisms of AF: Basic and Translational Science and Genetics

Attempts to develop safe and effective pharmacological therapy for AF have focused on atrium-selective drugs that take advantage of electrophysiological differences between the atrium and ventricle.[4–7] Heterogeneous abbreviation of the effective refractory period within the atrium provides the electric substrate for development of AF. The reduced effective refractory period results from abbreviation of the atrial action potential duration, which is caused by a decrease in the calcium channel current (I_{Ca}) and an increase in the potassium channel current (I_{K1}) and the constitutively active acetylcholine-sensitive current (CA I_{KACh}).[4–7] Maintenance of AF is facilitated by structural remodeling and additional abbreviation of the effective refractory period. The principal goal of pharmacological therapy is therefore to augment the effective refractory period.

Distinctions in the ion channel currents between the atrium and ventricle open the possibility for development of atrium-specific and -selective drugs for rhythm control of AF, which might avoid ventricular proarrhythmic effects. Atrium-specific targets include I_{Kur}, I_{KACh}, and the constitutively active I_{KACh}, the most investigated of which is inhibition of I_{Kur}.[7] Recent experimental studies have identified atrium-selective I_{Na} blockers that can effectively suppress AF while exerting little or no effect in the ventricles.[5,6] Combinations of

The American Heart Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

This document was approved by the American Heart Association Science Advisory and Coordinating Committee on May 9, 2011. A copy of the document is available at http://my.americanheart.org/statements by selecting either the "By Topic" link or the "By Publication Date" link. To purchase additional reprints, call 843-216-2533 or e-mail kelle.ramsay@wolterskluwer.com.

The American Heart Association requests that this document be cited as follows: Estes NAM 3rd, Sacco RL, Al-Khatib SM, Ellinor PT, Bezanson J, Alonso A, Antzelevitch C, Brockman RG, Chen P-S, Chugh SS, Curtis AB, DiMarco JP, Ellenbogen KA, Epstein AE, Ezekowitz MD, Fayad P, Gage BF, Go AS, Hlatky MA, Hylek EM, Jerosch-Herold M, Konstam MA, Lee R, Packer DL, Po SS, Prystowsky EN, Redline S, Rosenberg Y, Van Wagoner DR, Wood KA, Yue L, Benjamin EJ. American Heart Association Atrial Fibrillation Research Summit: a conference report from the American Heart Association. *Circulation*. 2011;124:363–372.

Expert peer review of AHA Scientific Statements is conducted at the AHA National Center. For more on AHA statements and guidelines development, visit http://my.americanheart.org/statements and select the "Policies and Development" link.

(*Circulation*. 2011;124:363-372.)

Circulation **is available at http://circ.ahajournals.org** **DOI: 10.1161/CIR.0b013e318224b037**

antiarrhythmic medications, such as amiodarone and ranolazine or dronedarone and ranolazine, are a promising avenue of investigation.[8,9] In animal models, such combinations produce potent synergistic effects that result in atrium-selective depression of sodium channel–dependent parameters and effective suppression of AF. Although synergism between atrium-selective drug combinations holds great promise, future work will be required to determine the safety and efficacy of such drug combinations in patients.

Multiple lines of evidence from basic, translational, and human studies suggest that atrial fibrosis plays an important role in the maintenance of AF.[10–13] Understanding of the mechanisms of atrial fibrogenesis will provide better targets for antifibrotic treatment of AF. There is a lack of detailed information about the biological and electrophysiological properties of fibroblasts and myofibroblasts under normal and AF conditions. Insights into signaling pathways suggest that atrial fibroblasts can be effective therapeutic targets for prevention of fibrosis.[14–16] Multiple agents with antifibrotic properties have demonstrated effects on reducing atrial fibrosis in animal models, but clinical evidence supporting their efficacy in AF prevention is currently lacking. Identification of fibroblast-specific genes will help to develop a fibroblast-specific knockout or transgenic mouse model.

Endothelin 1, a potent vasoconstrictor and mitogen involved in blood pressure regulation, may merit further investigation in AF because it modulates calcium cycling in cardiac myocytes and promotes fibroblast proliferation.[17] Because of the differential distribution of downstream signaling elements, endothelin 1 has a greater impact on atrial than on ventricular calcium cycling and contractility.[17] Interestingly, atrial endothelin 1 levels are elevated in the left atrium of patients with structural heart disease and persistent AF.[17]

Late gadolinium enhancement by cardiac magnetic resonance imaging is a highly specific and sensitive method for detecting scar in the ventricular myocardium.[18,19] Furthermore, it has been demonstrated recently that the same gadolinium-based extracellular contrast agents can be used for the quantification of extracellular remodeling and the detection of diffuse interstitial fibrosis.[20,21] Development of methods for direct assessment of myocardial collagen burden using extracellular collagen-binding contrast agents might allow imaging to characterize the presence and severity of atrial fibrosis noninvasively.

Over the past 5 years, family history has been established as a risk factor for AF.[22] Familial forms of AF have been described, and mutations have been identified in ion channel proteins and signaling molecules; however, these genes are rare causes of AF.[23–26] Genome-wide association studies have revealed genetic risk factors for AF unanticipated by prior knowledge. In one case, the genetic variants are adjacent to a transcription factor that specifies left atrial and pulmonary vein development. How variants at this and other loci lead to AF remains unclear. Future efforts should be focused on the identification of new genetic loci, determination of the mechanism by which polymorphisms are associated with the initiation or promotion of AF, and exploration of the relation between genetic data and clinical outcomes for AF. Systems biology approaches can provide important insights by addressing the complex interactions between risk factors and disease and thus may facilitate our understanding of novel AF mechanisms. The major recommendations related to basic and translational science and AF are summarized in Table 1.

Table 1. Recommendations From the AHA Atrial Fibrillation Research Summit: Mechanisms of AF: Basic and Translational Science and Genetics

Conduct research on the mechanisms and pathogenesis of AF.
Explore the relation between genetic data and clinical outcomes for AF.
Determine the properties of sodium channel blockers (kinetics and modes of binding and unbinding, lipid solubility, molecular size, and chemical structure) that influence atrial selectivity.
Define how electrical and structural remodeling and genetics modify atrial selectivity and the utility and safety of I_{Na} blockers for treatment of AF.
Define the short-term and long-term effects of atrium-selective I_{Na} blockers.
Determine methods for direct assessment of atrial myocardial collagen burden, including extracellular collagen-binding MRI contrast agents.
Evaluate the biological, electrophysiological, and signaling properties of fibroblasts and myofibroblasts under normal and AF conditions to identify novel targets for antifibrotic therapy.
Identify biological markers that will allow antifibrotic treatment at early stages of fibrogenesis.
Identify fibroblast-specific genes to allow development of a fibroblast-specific knockout transgenic mouse model.
Determine the role of endothelin 1 in AF.

AHA indicates American Heart Association; AF, atrial fibrillation; and MRI, magnetic resonance imaging.

Epidemiology, Outcomes, Cost, AF, and Stroke Prevention

The AHA strategic plan now sets a goal of improving the cardiovascular health of all Americans by 20% in 10 years.[27] The AHA plan emphasizes primordial prevention: preventing development of risk factors for cardiovascular disease and stroke.[27] Lifestyle and health factors, including blood pressure, weight, glucose, cholesterol, smoking, diet, and physical activity, collectively known as "Life's Simple Seven," are the focus of the strategic plan.[27] The prevalence of hypertension, obesity, and diabetes mellitus, 3 major risk factors for AF,[27–31] will be favorably influenced by successful lifestyle modifications. Because AF represents one of the most potent risk factors for stroke, prevention of hypertension, obesity, and diabetes mellitus should decrease the incidence of ischemic neurological events substantially.

Several factors are associated with electric and structural atrial remodeling of the left atrium and thus are potential targets for primary prevention of AF. These risk factors can be grouped into demographic, anthropometric, behavioral, and classic cardiovascular risk factors; cardiovascular disease; pulmonary disease; and hyperthyroidism. Additionally, multiple biomarkers are associated with AF, including inflammatory (C-reactive protein, interleukin 6), oxidative stress (nicotinamide adenine dinucleotide phosphate [NADPH] oxidase, natriuretic peptides), and vasoactive peptide (endothelin 1) markers.[32] The extent to which biomarkers are causally related to AF onset, as opposed to representing an epiphenomenon of cardiovascular remodeling, remains unknown.

Systems biology approaches may provide important insights for the prediction and prevention of AF because they address the complex interactions between risk factors and disease.[33] Already, genome-wide association studies have revealed genetic risk factors for AF that were unanticipated by prior knowledge.[23–26] Because of the modest effect sizes, replicating and validating such signals will necessitate the collaboration of multiple cohorts and studies.

Although risk-prediction models to identify individuals at increased risk of AF exist, several areas of uncertainty remain.[2] AF is more common in whites than in other racial groups,[26] but the mechanisms underlying these racial differences are unknown. Potential explanations include variation in diagnosis by race, presence of competing risks in minorities, different susceptibility to AF risk factors, and genetic determinants.[2,24–26] It is uncertain whether AF risk-prediction instruments developed in whites are generalizable to blacks or other ethnic groups. In addition, the proper translation of imaging tests, biomarkers, and genomic markers into clinical practice for improved risk prediction is uncertain. The development of genomic and clinical phenotypes that can distinguish different AF subtypes by functional burden, prognosis, and response to treatment would represent a fundamental advance. Classification into mechanistic and pathogenetic subtypes might allow targeted prevention and treatment in the future.

The outcomes of patients with AF can be improved by prompt diagnosis, appropriate treatment, adherence to practice guidelines, quality research that includes comparative effectiveness research, proper utilization of registries, and translation of research findings into better decision making. Evidence-based guidelines from the AHA/American Stroke Association emphasize the need for antithrombotic therapy, in particular warfarin, in eligible patients.[3,30] The efficacy of warfarin at reducing stroke in patients with AF has been well established by randomized clinical trials.[3] Indeed, the AHA Get With The Guidelines (GWTG) program has identified use of antithrombotic therapy among patients with AF as the key performance metric. Although more patients are being treated with antithrombotic therapy over time, community statistics continue to show large numbers of AF subjects who are not treated with antithrombotic therapy. Despite its demonstrated efficacy, warfarin use is associated with many challenges. Thus, novel vitamin K antagonists are in development. Dabigatran, a direct thrombin inhibitor studied in the RE-LY trial (Randomized Evaluation of Long-Term Anticoagulation Therapy), proved superior to warfarin without an increased incidence of major bleeding.[34] Several factor Xa inhibitors are being examined in phase 3 trials.[34–36] The comparative efficacy of these agents versus warfarin and their impact on utilization require further evaluation.

Data from large healthcare systems and electronic medical records may provide key data on practice patterns in AF management and contribute to improving the outcomes of patients with AF.[37] Such "real-world data" will facilitate comparative effectiveness research and may allow comparisons of the effectiveness of treatments as used in routine clinical practice. Given the important health policy implications of conclusions about the comparative effectiveness of various diagnostic and treatment approaches, caution should be exercised in drawing either positive or negative conclusions from observational analyses alone, although such findings may be instrumental in directing definitive prospective randomized trials that are sufficiently powered for noninferiority conclusions. The large numbers of patients in these databases also may allow examination of understudied subgroups. Several AF registries are being developed, including SAFARI (Safety of Atrial Fibrillation Ablation Registry Initiative), a national registry of AF catheter ablation procedures, and ORBIT-AF (Outcomes Registry for Better Informed Treatment of Atrial Fibrillation), whose primary objective is to assess real-world patterns of care and outcomes of patients with AF.[38] The major recommendations related to AF prevention and stroke prevention, epidemiology, outcomes, and cost are summarized in Table 2.

Table 2. Recommendations From the AHA Atrial Fibrillation Research Summit: Epidemiology, Outcomes, Cost, AF, and Stroke Prevention

Encourage international, multiinstitutional, multidisciplinary teams to collaborate in research and integrate approaches to AF prevention.
Determine the cause of variation in AF epidemiology, presentation, risk factors, and prognosis by age, sex, and race/ethnicity.
Develop strategies to detect and monitor AF more effectively in individuals and large populations.
Develop more accurate clinical, imaging, biomarker, and genomic phenotyping to more rigorously classify AF.
Evaluate AF risk-prediction instruments and the added clinical utility of novel markers.
Systematically ascertain the burden and impact of AF in all regions of the world.
Conduct randomized clinical trials of primordial, primary, and secondary AF prevention.
Refine stratification for risk of systemic thromboembolism, intracranial hemorrhage, and other major bleeding.
Improve implementation of proven stroke prevention guidelines, particularly in underserved populations.
Characterize the clinical outcomes, cost, and impact on quality of life and utilization of antithrombotic therapy of new anticoagulants.
Design, fund, and conduct rigorous comparative effectiveness and safety studies of AF therapeutic approaches.
Create and realign meaningful provider, institutional, and payer incentives to encourage participation in databases and registries.
Develop better models of cost in AF management that include quality of life, patient/provider adherence, clinical outcomes, and healthcare utilization.
Develop methods to facilitate the use of cost-effectiveness models by clinicians.

AHA indicates American Heart Association; AF, atrial fibrillation.

Meeting the Clinical Challenges and Redefining the Therapeutic Goals of AF

Antiarrhythmic drugs are likely to remain essential components of any comprehensive therapeutic strategy of maintaining sinus rhythm in patients with AF.[39] Recently, drug development for AF has emphasized multichannel blockers with less potential toxicity than amiodarone, including novel ion channel targets (eg, I_{Kur} blockers, late I_{Na} blockers, calcium current modulators), and non–ion-channel therapeutic targets (eg, fibrosis, gap junctions, and inflammation).[40,41]

Novel approaches being evaluated to prevent AF are the use of statins, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, and omega-3 fatty acids.[42] In addition to improving clinical outcomes in patients with heart failure and prior myocardial infarction, angiotensin-converting enzyme inhibitors and angiotensin receptor blocker agents appear to prevent the atrial electric and structural remodeling associated with AF through hemodynamic, antiproliferative, antiinflammatory, antioxidant, and antiapoptotic effects.[42] These agents have been shown to be antisympathetic and to prevent the development of left atrial stretch and interstitial fibrosis, as well as adverse atrial electric remodeling. However, these effects have not yet been clearly translated into clinical benefit, because there have been mixed results for the efficacy of angiotensin-converting enzyme inhibitors and angiotensin receptor blockers in prevention of AF. The evidence-based medicine that supports the use of these agents to prevent AF is insufficient to merit incorporating their use into clinical practice guidelines. Similarly, statins have been shown to possess antiinflammatory, antiischemic, antioxidant, and antiarrhythmic effects while having a beneficial effect on modulation of autonomic tone; however, there have been mixed results related to the efficacy of statins in the prevention of AF. Other potential targets for prevention of AF are aldosterone and endothelin 1 antagonists, but no meaningful clinical data are available.[42]

An important issue related to AF clinical trial design is the selection of optimal monitoring approaches and end point definitions.[39] Traditional end points for AF drug trials have been time to first symptomatic recurrence and total recurrences over time. The total burden of AF, however, may be a better measure of the effectiveness of therapy. Use of this end point will require better understanding of the advantages and disadvantages of intermittent monitoring, 30-day full-disclosure monitors, and implantable loop recorders to quantify the burden of AF and the correlation of AF burden with clinically important end points such as mortality, stroke, and quality of life. The prognostic significance of asymptomatic episodes of AF and the clinical implications of the burden of AF remain unknown.[39]

Although intermediate end points are more easily measured, adequate rate control and the number and timing of AF recurrences do not necessarily correspond to the clinical value of a drug and the patient's prognosis.[39,43,44] Recurrent AF as an outcome may be satisfactory in young patients with highly symptomatic AF; however, AF recurrence fails to tell us about the effects of a given therapy on the most important outcomes, including stroke incidence, cardiovascular morbidity, cardiovascular mortality, total mortality, healthcare costs, New York Heart Association functional class, quality of life, and exercise tolerance.

The ATHENA trial (A Placebo-Controlled, Double-Blind, Parallel Arm Trial to Assess the Efficacy of Dronedarone 400 mg BID for the Prevention of Cardiovascular Hospitalization or Death From Any Cause in Patients With Atrial Fibrillation/Atrial Flutter) showed that antiarrhythmic drug therapy decreases cardiovascular hospitalizations.[45] The major mechanism of dronedarone benefit is probably linked to a combination of decreasing the recurrence of AF and decreasing heart rate, particularly the ventricular response in the presence of AF.[45] However, other effects that antiarrhythmic drugs may have on cardiovascular and noncardiovascular mortality or morbidity, as well as extracardiac effects, remain uncertain.[46] Assessment of rhythm control should be focused not only on recurrence of symptomatic arrhythmias but also on objective measurements of cardiovascular morbidity, mortality, hospitalizations, functional status, quality of life, and cost.[39]

There remain many limitations to the randomized controlled trials evaluating catheter ablation of AF compared with antiarrhythmic agents.[47,48] Further trials are needed to evaluate catheter ablation of AF versus the strategy of rate control or rhythm control. Mortality, freedom from recurrent AF, costs, resource utilization, cost-effectiveness, and impact on quality of life should be included as end points. The Catheter Ablation versus Antiarrhythmic Drug Therapy for Atrial Fibrillation Trial (CABANA) is an ongoing multicenter study of catheter ablation versus pharmacological therapy as treatment of symptomatic AF.[49] This multicenter randomized study is intended to address many of the limitations of previous studies. Nonetheless, like other clinical trials, CABANA will provide only limited insight into important aspects of this procedure such as the techniques, safety, and long-term effectiveness of AF ablation in routine clinical practice. There is a need to address the limitations of available data and other key concerns about the safety of AF ablation with robust registry data.[38] As noted for catheter ablation of AF, complications, mortality, freedom from recurrent AF, costs, resource utilization, cost-effectiveness, and impact on quality of life should be included as end points for the maze procedure and other surgical approaches to AF.[48,50,51] Additionally, prospective multicenter clinical trials are needed to better define the relative safety and efficacy of various surgical tools and techniques for surgical ablation of AF.[48,50,51]

The investigation of advanced approaches to AF management requires an understanding of the complex interplay between AF and heart failure. The prevalences of both conditions are increasing in concert, particularly with an aging population.[52] Heart failure and AF frequently coexist, with heart failure representing an important risk factor for developing AF and AF contributing importantly to the morbidity of patients with heart failure, particularly those without left ventricular dilation and with preserved left ventricular ejection fractions.[1,2] Both the hemodynamic and neurohormonal perturbations of heart failure likely contribute to the pathogenesis of AF via both mechanical and structural changes within the atria. A number of studies have suggested that inhibition of the renin-angiotensin-aldosterone axis reduces the incidence of AF, an effect that may be mediated via reduced atrial stretch, prevention of adverse atrial remodeling, or both.[42] Research into prevention and management of AF should consider the interaction of these 2 conditions, exploration of mechanistic interactions, examination of relationships between treatment effects on both arrhythmic and heart failure–related end points, and addressing the expanding population in which these 2 conditions coexist.

A substantial body of clinical and experimental data indicates that sleep apnea and AF commonly coaggregate.[53–56] Estimates from population-based research suggest that individuals with moderate to severe sleep apnea are

Table 3. Recommendations From the AHA Atrial Fibrillation Research Summit Meeting: Meeting the Clinical Challenges and Redefining the Therapeutic Goals in AF

Evaluate the efficacy and safety of agents with antifibrotic properties.
Evaluate alternative upstream therapies (statins, ACE inhibitors/ARBs, fish oil) to prevent onset and recurrences of AF.
Evaluate efficacy and safety of atrium-selective I_{Na} blocker combinations shown to exert potent synergistic actions in experimental models of AF (ie, ranolazine and dronedarone combination).
Identify appropriate end points for observational studies, comparative effectiveness studies, and clinical trials of AF.
Evaluate whether the burden of AF and benefits of therapies are related to symptoms, QOL, functional status, and cardiovascular outcomes.
Develop a QOL metric that is specific to patients with AF and one that can measure the effect of AF on symptoms as well as outcomes.
Develop methods to efficiently measure AF burden, outcomes, and cost rather than process metrics.
Explore the interactions between AF and heart failure, including mechanistic interactions and relationships between treatment effects on both arrhythmic and heart failure–related end points; develop management strategies that address the expanding population in which these 2 conditions coexist.
Identify the heart rates and duration of AF that result in tachycardia-induced cardiomyopathy.
Compare rhythm control by use of antiarrhythmic drug therapy with catheter ablation of AF.
Conduct clinical and mechanistic studies defining the interactions among sleep apnea–mediated changes in cardiac structure, autonomic function, and inflammation and their impact on AF.
Define strategies for prevention of postoperative AF.
Systematically evaluate strategies for a surgical cure of AF.
Provide both provider and health system incentives to design and embed quality and performance improvement in EMRs, track application of evidence-based therapies and improve patient outcomes.
Design and research decision support tools to enhance clinicians' abilities to implement evidence-based treatment and guidelines.

AHA indicates American Heart Association; AF, atrial fibrillation; ACE, angiotensin-converting enzyme; ARBs, angiotensin receptor blockers; QOL, quality of life; and EMRs, electronic medical records.

approximately 4-fold more likely to have AF than those without apnea and that the occurrence of apneas during sleep may serve as a trigger of paroxysms of AF.[57–59] These observations suggest that sleep apnea contributes to AF incidence and recurrence or that there are common mechanisms, such as altered autonomic tone, linked to the pathogenesis of both conditions. Because sleep apnea prevalence appears to be increasing in conjunction with the obesity epidemic, it is possible that unrecognized and untreated sleep apnea may be a significant contributor to the rising population burden of AF. Sleep apnea may provide a new intervention target for the prevention and management of AF. Reversal of sleep apnea–associated hypoxemia, intrathoracic swings, and autonomic imbalance may attenuate triggers for AF.

Multifaceted strategies to facilitate the process of improving clinical care have emerged, with an emphasis on evidence-based medicine, clinical practice guidelines, quality metrics and performance measures, and patient outcomes.[60] By facilitating measurements of cardiovascular healthcare quality, performance measurement sets may serve as vehicles to accelerate appropriate translation of scientific evidence into clinical practice. Application of performance measures related to AF should provide a mechanism through which the quality of medical care can be measured to improve patient outcomes. The major clinical recommendations related to AF are summarized in Table 3.

Conclusions

Although considerable progress has been made in understanding the mechanisms of AF and preventive and treatment strategies, it is evident that much remains unknown. Prevention and treatment of AF will ultimately depend on understanding the pathophysiology in the individual patient. Understanding of several key aspects of AF, including the patterns of its occurrence in different populations, risk factors, and underlying pathophysiology, will result in identification and testing of prevention and treatment strategies. A concerted effort is thus needed on several fronts, as outlined by the recommendations in Tables 1, 2, and 3, to optimally predict, prevent, and treat AF and thereby improve patient outcomes.

Sources of Funding

American Heart Association's Atrial Fibrillation Research Summit was sponsored by Sanofi US.

Appendix

Session I: Epidemiology, Outcomes, Cost, AF, and Stroke Prevention

Section Editor: Sana M. Al-Khatib, MD, Duke University, Durham, NC. *Moderators:* Mark A. Hlatky, MD, Stanford School of Medicine, Stanford, CA; Elaine Hylek, MD, MPH, Massachusetts General Hospital, Boston, MA.

Translating Evidence Into Practice: Reducing Death and Disability Due to Stroke—In Perspective of Ideal Health and Prevention: Ralph L. Sacco, MS, MD, FAHA, FAAN, University of Miami, Miami, FL.

Prediction and Prevention of AF: The Decade Ahead: Emelia J. Benjamin, MD, ScM, FAHA, Boston University Schools of Medicine and Public Health, Boston, MA.

The Global Burden of AF: Sumeet S. Chugh, MD, Cedars-Sinai Medical Center, Los Angeles, CA.

AF in Ethnic Minorities: What Are the Unanswered Questions? Alvaro Alonso, MD, MPH, PhD, University of Minnesota, Minneapolis, MN.

Utilizing Health Maintenance Organizations and Electronic Medical Records to Advance Knowledge in AF: Alan S. Go, MD, Kaiser Permanente of Northern California, Oakland, CA.

Improving Patient Outcomes in AF: Registries, Comparative Effectiveness: Sana M. Al-Khatib, MD, Duke University, Durham, NC.

Developing Better Models of Cost in AF Management: Brian F. Gage, MD, Washington University, St Louis, MO.

New Developments in Anticoagulation for AF: Michael D. Ezekowitz, MBChB, DPhil, FRCP, FAHA, MD, Lankenau Institute for Medical Research, Wynnewood, PA.

Session II: Mechanisms of AF: Basic and Translational Science and Genetics

Section Editor: Patrick Ellinor, MD, PhD. *Moderators:* Sunny S. Po, MD, PhD, University of Oklahoma Health Sciences Center, Oklahoma City, OK; Patrick T. Ellinor, MD, PhD, Massachusetts General Hospital, Boston, MA.

Atrioventricular Distinctions in Ion Channels as the Basis for Development of Atrial-Selective Drugs in the Management of AF: Charles Antzelevitch, PhD, FACC, FAHA, FHRS, Masonic Medical Research Laboratory, Utica, NY.

Sleep Apnea, Experimental Hypertension, and AF: Autonomic and Structural Considerations: Susan Redline, MD, MPH, Brigham and Women's Hospital, Boston, MA.

Novel MRI Markers of Cardiac Fibrosis: Michael Jerosch-Herold, PhD, Brigham and Women's Hospital, Boston, MA.

Fibrosis and AF: Lixia Yue, PhD, University of Connecticut Health Center, Farmington, CT.

Mechanisms of AF in Hypertrophy, Endothelin, Angiotensin Signaling: David R. Van Wagoner, PhD, FAHA, Cleveland Clinic, Cleveland, OH.

Novel Pathways for Pharmacological Treatment and Prevention of AF: Peng-Sheng Chen, MD, Indiana University School of Medicine, Indianapolis, IN.

Session III: Meeting the Clinical Challenges in AF

Section Editor: N.A. Mark Estes III, MD. *Moderators:* Kathryn Wood, PhD, RN, Duke University, Durham, NC; A.E. Epstein, MD, FAHA, FACC, FHRS, VA Medical Center, Philadelphia, PA; Marv Konstam, MD, Tufts University, Boston, MA.

Therapies to Prevent AF: Statins, ACE-I/ARB, Omega-3 Fatty Acids: Anne B. Curtis, MD, FHRS, FACC, FAHA, The State University of New York, University at Buffalo.

Strategies for AF Management: Reevaluating Rate and Rhythm Control: Ken A. Ellenbogen, MD, FAHA, Virginia Commonwealth University, Richmond, VA.

New Pharmacological Agents for AF: John P. DiMarco, MD, PhD, University of Virginia Health System, Charlottesville, VA.

Ablation of AF: Addressing the Gaps in Knowledge: Douglas L. Packer, MD, Mayo Clinic, Rochester, MN.

Session IV: Redefining the Therapeutic Goals of AF

New Initiatives From the NHLBI: Yves D. Rosenberg, MD, MPH, National Heart, Lung, and Blood Institute, Bethesda, MD.

Prevention of Postoperative AF and Advancing the Surgical Frontier for Cure of AF: Richard Lee, MD, Northwestern Medical Center, Chicago, IL.

Developing and Validating Standardized and Quantitative Approaches to Understanding the Burden of Asymptomatic and Symptomatic AF on Quality of Life: Eric N. Prystowsky, MD, FAHA, St Vincent Hospital, Indianapolis, IN.

AF Clinical Guidelines, Performance Measures, and Quality Metrics: Improving Clinical Outcomes in AF Patients: N.A. Mark Estes III, MD, FAHA, Tufts University, Boston, MA.

Disclosures

Advisory Group Disclosures

Advisory Group Member	Employment	Research Grant	Other Research Support	Speakers' Bureau/ Honoraria	Expert Witness	Ownership Interest	Consultant/Advisory Board	Other
N.A. Mark Estes III	Pratt Medical Group, Tufts Medical Center	None	None	Boston Scientific*	None	None	I received payment (totaling <$10 000) for serving as Chair of an Advisory Board developing CME-approved educational content for healthcare professionals and patients related to atrial fibrillation. The Advisory Board consisted of academic cardiologists, neurologists, and an internist. The funding was from an unrestricted educational grant from Boehringer Ingelheim. The CME was provided by the University of Nebraska School of Medicine.*	None
Sana M. Al-Khatib	Duke University Hospital	Bristol-Myers Squibb†; Medtronic†	None	Medtronic*	None	None	None	None
Alvaro Alonso	University of Minnesota	AHA†; NIH†	None	None	None	None	None	None
Charles Antzelevitch	Masonic Medical Research Laboratory	AstraZeneca*; Cardiome†; Gilead Sciences†; Lundbeck*; Merck†; Solvay†	None	None	None	None	AstraZeneca*; Gilead Sciences†; Lundbeck*	None
Emelia J. Benjamin	Boston University School of Medicine	None	None	None	None	None	None	None
Judy Bezanson	American Heart Association	None	None	None	None	None	None	None
Randall G. Brockman	FDA	None	None	None	None	None	None	None
Peng-Sheng Chen	Indiana University	Medtronic†; St Jude†	None	None	None	None	None	None

(*Continued*)

Advisory Group Disclosures, *Continued*

Advisory Group Member	Employment	Research Grant	Other Research Support	Speakers' Bureau/ Honoraria	Expert Witness	Ownership Interest	Consultant/Advisory Board	Other
Sumeet S. Chugh	Cedars-Sinai Medical Center	NHLBI†	None	None	None	None	None	None
Anne B. Curtis	University of South Florida	Medtronic*; St Jude Medical*	None	Medtronic*; Sanofi-Aventis†; St Jude Medical*	None	None	Biosense Webster*; Medtronic†; Sanofi-Aventis*; St Jude Medical*	None
John P. DiMarco	University of Virginia	St Jude Medical*	None	None	None	None	Astellas*; Sanofi-Aventis*; Medtronic†; St Jude Medical*	None
Kenneth A. Ellenbogen	VCU School of Medicine	Biosense Webster†; Boston Scientific†; Medtronic†; Sanofi-Aventis*; Spectranetics†; St Jude Medical†	None	Biotouch*; Boston Scientific†; Medtronic*; Sanofi-Aventis*; St Jude Medical†	Boston Scientific†; Medtronic*;	None	Biotronik†; Biosense Webster*; Boston Scientific†; CardioNet†; EBR†; Medtronic*; Sanofi-Aventis†; St Jude Medical†	None
Patrick T. Ellinor	Massachusetts General Hospital	NIH†	None	None	None	None	Sanofi-Aventis*	None
Andrew E. Epstein	Columbia University	None	None	None	None	None	None	None
Michael D. Ezekowitz	Lankenau Institute for Medical Research	ARYx Therapeutics†; Boehringer Ingelheim†; Daiichi Sankyo*; PORTOLA†	None	None	None	None	ARYx Therapeutics†; AstraZeneca*; Boehringer Ingelheim*; Bristol-Myers Squibb*; Daiichi Sankyo*; Medtronic*; PORTOLA†; Sanofi*; Wyeth*	None
Pierre Fayad	University of Nebraska Medical Center	None	None	None	None	None	None	None
Brian F. Gage	Washington University	None	None	None	None	None	None	None
Alan S. Go	Kaiser Permanente	None	None	None	None	None	None	None
Mark A. Hlatky	Standford University	None	None	None	None	None	None	None
Elaine M. Hylek	Boston Medical Center	None	None	None	None	None	None	None
Michael Jerosch-Herold	Brigham & Women's Hospital	None	None	None	None	None	None	None
Marvin A. Konstam	Tufts Medical Center	None	None	None	None	None	None	None
Richard Lee	Northwestern University	Medtronic*	None	None	None	None	Medtronic*	None
Douglas L. Packer	Mayo Clinic	Biosense Webster*; Boston Scientific*; CryoCath*; EP Limited*; Medtronic*; Minnesota Partnership for Biotechnology*; NIH*; Siemens AG*	None	None	None	Financial interest in mapping technology licensed to St Jude Medical†	Biosense Webster*; Boston Scientific*; CyberHeart*; Medtronic*; Sanofi-Aventis*; St Jude Medical*; Toray Industries*	None
Sunny S. Po	University of Oklahoma Health Sciences Center	None	None	Biosense*	None	None	None	None

(*Continued*)

Advisory Group Disclosures, *Continued*

Advisory Group Member	Employment	Research Grant	Other Research Support	Speakers' Bureau/ Honoraria	Expert Witness	Ownership Interest	Consultant/Advisory Board	Other
Eric N. Prystowsky	The Care Group	Sanofi-Aventis†; fellowship support from Boston Scientific†, Medtronic†, and St Jude†	None	None	None	CardioNet†; Stereotaxis†	Boehringer Ingelheim†; Medtronic†; Sanofi-Aventis†; Stereotaxis†	None
Susan Redline	Case Western Reserve University/ Brigham and Women's Hospital	Received contract from Dymedix Inc to perform apnea/hypopnea sensor validation studies; receive CPAP units by Philips Respironics for use in NIH-funded research	None	None	None	None	None	None
Yves Rosenberg	National Institutes of Health	None	None	None	None	None	None	None
Ralph L. Sacco	University of Miami	None	None	None	None	None	None	None
David R. Van Wagoner	Cleveland Clinic	GlaxoSmithKline†	None	None	None	None	None	None
Kathryn A. Wood	Duke University School of Nursing	None	None	None	None	None	None	None
Lixia Yue	University of Connecticut Health Center	NIH†	None	None	None	None	None	None

This table represents the relationships of advisory group members that may be perceived as actual or reasonably perceived conflicts of interest as reported on the Disclosure Questionnaire, which all members of the writing group are required to complete and submit. A relationship is considered to be "significant" if (1) the person receives $10 000 or more during any 12-month period, or 5% or more of the person's gross income; or (2) the person owns 5% or more of the voting stock or share of the entity, or owns $10 000 or more of the fair market value of the entity. A relationship is considered to be "modest" if it is less than "significant" under the preceding definition.

*Modest.

†Significant.

References

1. Benjamin EJ, Chen PS, Bild DE, Mascette AM, Albert CM, Alonso A, Calkins H, Connolly SJ, Curtis AB, Darbar D, Ellinor PT, Go AS, Goldschlager NF, Heckbert SR, Jalife J, Kerr CR, Levy D, Lloyd-Jones DM, Massie BM, Nattel S, Olgin JE, Packer DL, Po SS, Tsang TS, Van Wagoner DR, Waldo AL, Wyse DG. Prevention of atrial fibrillation: report from a National Heart, Lung, and Blood Institute Workshop. *Circulation*. 2009;119:606–618.
2. Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB Sr, Newton-Cheh C, Yamamoto JF, Magnani JW, Tadros TM, Kannel WB, Wang TJ, Ellinor PT, Wolf PA, Vasan RS, Benjamin EJ. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *Lancet*. 2009;373:739–745.
3. Fuster V, Rydén LE, Cannom DS, Crijns HJ, Curtis AB, Ellenbogen KA, Halperin JL, Le Heuzey JY, Kay GN, Lowe JE, Olsson SB, Prystowsky EN, Tamargo JL, Wann S, Smith SC Jr, Jacobs AK, Adams CD, Anderson JL, Antman EM, Halperin JL, Hunt SA, Nishimura R, Ornato JP, Page RL, Riegel B, Priori SG, Blanc JJ, Budaj A, Camm AJ, Dean V, Deckers JW, Despres C, Dickstein K, Lekakis J, McGregor K, Metra M, Morais J, Osterspey A, Tamargo JL, Zamorano JL; American College of Cardiology/American Heart Association Task Force on Practice Guidelines; European Society of Cardiology Committee for Practice Guidelines; European Heart Rhythm Association; Heart Rhythm Society. ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Revise the 2001 Guidelines for the Management of Patients With Atrial Fibrillation): developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society [published correction appears in *Circulation*. 2007;116:e138]. *Circulation*. 2006;114:e257–e354.
4. Li X, Zima AV, Sheikh F, Blatter LA, Chen J. Endothelin-1–induced arrhythmogenic Ca^{2+} signaling is abolished in atrial myocytes of inositol-1,4,5-trisphosphate(IP3)-receptor type 2–deficient mice. *Circ Res*. 2005;96:1274–1281.
5. Burashnikov A, Di Diego JM, Zygmunt AC, Belardinelli L, Antzelevitch C. Atrium-selective sodium channel block as a strategy for suppression of atrial fibrillation: differences in sodium channel inactivation between atria and ventricles and the role of ranolazine. *Circulation*. 2007;116:1449–1457.
6. Ehrlich JR, Nattel S. Atrial-selective pharmacological therapy for atrial fibrillation: hype or hope? *Curr Opin Cardiol*. 2009;24:50–55.
7. Burashnikov A, Antzelevitch C. Can inhibition of I_{Kur} promote atrial fibrillation? *Heart Rhythm*. 2008;5:1304–1309.
8. Burashnikov A, Di Diego JM, Sicouri S, Ferreiro M, Carlsson L, Antzelevitch C. Atrial-selective effects of chronic amiodarone in the management of atrial fibrillation. *Heart Rhythm*. 2008;5:1735–1742.
9. Burashnikov A, Sicouri S, Di Diego JM, Belardinelli L, Antzelevitch C. Synergistic effect of the combination of ranolazine and dronedarone to suppress atrial fibrillation. *J Am Coll Cardiol*. 2010;56:1216–1224.
10. Burstein B, Nattel S. Atrial fibrosis: mechanisms and clinical relevance in atrial fibrillation. *J Am Coll Cardiol*. 2008;51:802–809.

11. Frustaci A, Chimenti C, Bellocci F, Morgante E, Russo MA, Maseri A. Histological substrate of atrial biopsies in patients with lone atrial fibrillation. *Circulation*. 1997;96:1180–1184.
12. Verheule S, Sato T, Everett T 4th, Engle SK, Otten D, Rubart-von der Lohe M, Nakajima HO, Nakajima H, Field LJ, Olgin JE. Increased vulnerability to atrial fibrillation in transgenic mice with selective atrial fibrosis caused by overexpression of TGF-beta1. *Circ Res*. 2004;94:1458–1465.
13. Xiao HD, Fuchs S, Campbell DJ, Lewis W, Dudley SC Jr, Kasi VS, Hoit BD, Keshelava G, Zhao H, Capecchi MR, Bernstein KE. Mice with cardiac-restricted angiotensin-converting enzyme (ACE) have atrial enlargement, cardiac arrhythmia, and sudden death. *Am J Pathol*. 2004;165:1019–1032.
14. Baudino TA, McFadden A, Fix C, Hastings J, Price R, Borg TK. Cell patterning: interaction of cardiac myocytes and fibroblasts in three-dimensional culture. *Microsc Microanal*. 2008;14:117–125.
15. Du J, Xie J, Zhang Z, Tsujikawa H, Fusco D, Silverman D, Liang B, Yue L. TRPM7-mediated Ca^{2+} signals confer fibrogenesis in human atrial fibrillation. *Circ Res*. 2010;106:992–1003.
16. Li GR, Wang HB, Qin GW, Jin MW, Tang Q, Sun HY, Du XL, Deng XL, Zhang XH, Chen JB, Chen L, Xu XH, Cheng LC, Chiu SW, Tse HF, Vanhoutte PM, Lau CP. Acacetin, a natural flavone, selectively inhibits human atrial repolarization potassium currents and prevents atrial fibrillation in dogs. *Circulation*. 2008;117:2449–2457.
17. Mayyas F, Niebauer M, Zurick A, Barnard J, Gillinov AM, Chung MK, Van Wagoner DR. Association of left atrial endothelin 1 with atrial rhythm, size and fibrosis in patients with structural heart disease. *Circ Arrhythm Electrophysiol*. 2010;3:369–379.
18. Kim RJ, Wu E, Rafael A, Chen E-L, Parker MA, Simonetti O, Klocke FJ, Bonow RO, Judd RM. The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction. *N Engl J Med*. 2000;343:1445–1453.
19. Kim RJ, Fieno DS, Parrish TB, Harris K, Chen EL, Simonetti O, Bundy J, Finn JP, Klocke FJ, Judd RM. Relationship of MRI delayed contrast enhancement to irreversible injury, infarct age, and contractile function. *Circulation*. 1999;100:1992–2002.
20. Flett AS, Hayward MP, Ashworth MT, Hansen MS, Taylor AM, Elliott PM, McGregor C, Moon JC. Equilibrium contrast cardiovascular magnetic resonance for the measurement of diffuse myocardial fibrosis: preliminary validation in humans. *Circulation*. 2010;122:138–144.
21. Broberg CS, Chugh S, Conklin C, Sahn DJ, Jerosch-Herold M. Quantification of diffuse myocardial fibrosis and its association with myocardial dysfunction in congenital heart disease. *Circ Cardiovasc Imaging*. 2010;3:727–734.
22. Fox CS, Parise H, D'Agostino RB Sr, Lloyd-Jones DM, Vasan RS, Wang TJ, Levy D, Wolf PA, Benjamin EJ. Parental atrial fibrillation as a risk factor for atrial fibrillation in offspring. *JAMA*. 2004;291:2851–2855.
23. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, Jonasdottir A, Baker A, Thorleifsson G, Kristjansson K, Palsson A, Blondal T, Sulem P, Backman VM, Hardarson GA, Palsdottir E, Helgason A, Sigurjonsdottir R, Sverrisson JT, Kostulas K, Ng MC, Baum L, So WY, Wong KS, Chan JC, Furie KL, Greenberg SM, Sale M, Kelly P, MacRae CA, Smith EE, Rosand J, Hillert J, Ma RC, Ellinor PT, Thorgeirsson G, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature*. 2007;448:353–357.
24. Benjamin EJ, Rice KM, Arking DE, Pfeufer A, van Noord C, Smith AV, Schnabel RB, Bis JC, Boerwinkle E, Sinner MF, Dehghan A, Lubitz SA, D'Agostino RB Sr, Lumley T, Ehret GB, Heeringa J, Aspelund T, Newton-Cheh C, Larson MG, Marciante KD, Soliman EZ, Rivadeneira F, Wang TJ, Eiríksdottir G, Levy D, Psaty BM, Li M, Chamberlain AM, Hofman A, Vasan RS, Harris TB, Rotter JI, Kao WH, Agarwal SK, Stricker BH, Wang K, Launer LJ, Smith NL, Chakravarti A, Uitterlinden AG, Wolf PA, Sotoodehnia N, Köttgen A, van Duijn CM, Meitinger T, Mueller M, Perz S, Steinbeck G, Wichmann HE, Lunetta KL, Heckbert SR, Gudnason V, Alonso A, Kääb S, Ellinor PT, Witteman JC. Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry. *Nat Genet*. 2009;41:879–881.
25. Ellinor PT, Lunetta KL, Glazer NL, Pfeufer A, Alonso A, Chung MK, Sinner MF, de Bakker PI, Mueller M, Lubitz SA, Fox E, Darbar D, Smith NL, Smith JD, Schnabel RB, Soliman EZ, Rice KM, Van Wagoner DR, Beckmann BM, van Noord C, Wang K, Ehret GB, Rotter JI, Hazen SL, Steinbeck G, Smith AV, Launer LJ, Harris TB, Makino S, Nelis M, Milan DJ, Perz S, Esko T, Köttgen A, Moebus S, Newton-Cheh C, Li M, Möhlenkamp S, Wang TJ, Kao WH, Vasan RS, Nöthen MM, MacRae CA, Stricker BH, Hofman A, Uitterlinden AG, Levy D, Boerwinkle E, Metspalu A, Topol EJ, Chakravarti A, Gudnason V, Psaty BM, Roden DM, Meitinger T, Wichmann HE, Witteman JC, Barnard J, Arking DE, Benjamin EJ, Heckbert SR, Kääb S. Common variants in KCNN3 are associated with lone atrial fibrillation. *Nat Genet*. 2010;42:240–244.
26. Alonso A, Agarwal SK, Soliman EZ, Ambrose M, Chamberlain AM, Prineas RJ, Folsom AR. Incidence of atrial fibrillation in whites and African Americans: the Atherosclerosis Risk in Communities (ARIC) study. *Am Heart J*. 2009;158:111–117.
27. Lloyd Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, Greenlund K, Daniels S, Nichol G, Tomaselli GF, Arnett DK, Fonarow GC, Ho PM, Lauer MS, Masoudi FA, Robertson RM, Roger V, Schwamm LH, Sorlie P, Yancy CW, Rosamond WD; American Heart Association Strategic Planning Task Force and Statistics Committee. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation*. 2010;121:586–613.
28. Ardestani A, Hoffman HJ, Cooper HA. Obesity and Outcomes among patients with established atrial fibrillation. *Am J Cardiol*. 2010;106:369–373.
29. Goldstein LB, Adams R, Alberts MJ, Appel LJ, Brass LM, Bushnell CD, Culebras A, Degraba TJ, Gorelick PB, Guyton JR, Hart RG, Howard G, Kelly-Hayes M, Nixon JV, Sacco RL. Primary prevention of ischemic stroke: a guideline from the American Heart Association/American Stroke Association Stroke Council: cosponsored by the Atherosclerotic Peripheral Vascular Disease Interdisciplinary Working Group; Cardiovascular Nursing Council; Clinical Cardiology Council; Nutrition, Physical Activity, and Metabolism Council; and the Quality of Care and Outcomes Research Interdisciplinary Working Group [published correction appears in *Stroke*. 2007;38:207]. *Stroke*. 2006;37:1583–1633.
30. Schwamm L, Fayad P, Acker JE 3rd, Duncan P, Fonarow GC, Girgus M, Goldstein LB, Gregory T, Kelly Hayes M, Sacco RL, Saver JL, Segrest W, Solis P, Yancy CW. Translating evidence into practice: a decade of efforts by the American Heart Association/American Stroke Association to reduce death and disability due to stroke: a presidential advisory from the American Heart Association/American Stroke Association. *Stroke*. 2010;41:1051–1065.
31. Sacco RL, Adams R, Albers G, Alberts MJ, Benavente O, Furie K, Goldstein LB, Gorelick P, Halperin J, Harbaugh R, Johnston SC, Katzan I, Kelly-Hayes M, Kenton EJ, Marks M, Schwamm LH, Tomsick T. Guidelines for prevention of stroke in patients with ischemic stroke or transient ischemic attack: a statement for healthcare professionals from the American Heart Association/American Stroke Association Council on Stroke: co-sponsored by the Council on Cardiovascular Radiology and Intervention. *Stroke*. 2006;37:577–617.
32. Issac TT, Dokainish H, Lakkis NM. Role of inflammation in initiation and perpetuation of atrial fibrillation: a systematic review of the published data. *J Am Coll Cardiol*. 2007;50:2021–2028.
33. Lusis AJ, Weiss JN. Cardiovascular networks: systems-based approaches to cardiovascular disease. *Circulation*. 2010;121:157–170.
34. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, Pogue J, Reilly PA, Themeles E, Varrone J, Wang S, Alings M, Xavier D, Zhu J, Diaz R, Lewis BS, Darius H, Diener HC, Joyner CD, Wallentin L; RE-LY Steering Committee and Investigators. Dabigatran versus warfarin in patients with atrial fibrillation [published correction appears in *N Engl J Med*. 2010;363:1877]. *N Engl J Med*. 2009;361:1139–1151.
35. Lopes RD, Alexander JH, Al-Khatib SM, Ansell J, Diaz R, Easton JD, Gersh BJ, Granger CB, Hanna M, Horowitz J, Hylek EM, McMurray JJ, Verheugt FW, Wallentin L; ARISTOTLE Investigators. Apixaban for Reduction In STroke and Other ThromboembоLic Events in atrial fibrillation (ARISTOTLE) trial: design and rationale [published correction appears in *Am Heart J*. 2010;159:1162]. *Am Heart J*. 2010;159:331–339.
36. ROCKET AF Study Investigators. Rivaroxaban-once daily, oral, direct factor Xa inhibition compared with vitamin K antagonism for prevention of stroke and Embolism Trial in Atrial Fibrillation: rationale and design of the ROCKET AF study. *Am Heart J*. 2010;159:340–347.
37. Go AS, Magid DJ, Wells B, Sung SH, Cassidy-Bushrow AE, Greenlee RT, Langer RD, Lieu TA, Margolis KL, Masoudi FA, McNeal CJ, Murata GH, Newton KM, Novotny R, Reynolds K, Roblin DW, Smith DH, Vupputuri S, White RE, Olson J, Rumsfeld JS, Gurwitz JH. The Cardiovascular Research Network: a new paradigm for cardiovascular quality and outcomes research. *Circ Cardiovasc Qual Outcomes*. 2008;1:138–147.
38. Al-Khatib SM, Calkins H, Eloff BC, Packer DL, Ellenbogen KA, Hammill SC, Natale A, Page RL, Prystowsky E, Jackman WM, Ste-

venson WG, Waldo AL, Wilber D, Kowey P, Yaross MS, Mark DB, Reiffel J, Finkle JK, Marinac-Dabic D, Pinnow E, Sager P, Sedrakyan A, Canos D, Gross T, Berliner E, Krucoff MW. Planning the Safety of Atrial Fibrillation Ablation Registry Initiative (SAFARI) as a collaborative pan-stakeholder critical path registry model: a Cardiac Safety Research Consortium "Incubator" Think Tank. *Am Heart J*. 2010;159:17–24.

39. Prystowsky EN, Camm J, Lip GY, Allessie M, Bergmann JF, Breithardt G, Brugada J, Crijns H, Ellinor PT, Mark D, Naccarelli G, Packer D, Tamargo J. The impact of new and emerging clinical data on treatment strategies for atrial fibrillation. *J Cardiovasc Electrophysiol*. 2010;21:946–958.
40. Page RL, Roden DM. Drug therapy for atrial fibrillation: where do we go from here? *Nat Rev Drug Discov*. 2005;4:899–910.
41. Mason PK, DiMarco JP. New pharmacologic agents for arrhythmias. *Circ Arrhythmia Electrophysiol*. 2009;2:588–597.
42. Guglin M, Garcia M, Yarnoz MJ, Curtis AB. Non-antiarrhythmic medications for atrial fibrillation: from bench to clinical practice. *J Interv Card Electrophysiol*. 2008;22:119–128.
43. Wyse DG, Waldo AL, DiMarco JP, Domanski MJ, Rosenberg Y, Schron EB, Kellen JC, Greene HL, Mickel MC, Dalquist JE, Corley SD; Atrial Fibrillation Follow-up Investigation of Rhythm Management (AFFIRM) Investigators. A comparison of rate control and rhythm control in patients with atrial fibrillation. *N Engl J Med*. 2002;347:1825–1833.
44. Van Gelder IC, Groenveld HF, Crijns HJ, Tuininga YS, Tijssen JG, Alings AM, Hillege HL, Bergsma-Kadijk JA, Cornel JH, Kamp O, Tukkie R, Bosker HA, Van Veldhuisen DJ, Van den Berg MP; RACE II Investigators. Lenient versus strict rate control in patients with atrial fibrillation. *N Engl J Med*. 2010;362:1363–1373.
45. Hohnloser SH, Crijns HJ, van Eickels M, Gaudin C, Page RL, Torp-Pedersen C, Connolly SJ; ATHENA Investigators. Effect of dronedarone on cardiovascular events in atrial fibrillation [published corrections appear in *N Engl J Med*. 2009;360:2487 and *N Engl J Med*. 2011;364:1481]. *N Engl J Med*. 2009;360:668–678.
46. Connolly SJ, Crijns HJ, Torp-Pedersen C, van Eickels M, Gaudin C, Page RL, Hohnloser SH; ATHENA Investigators. Analysis of stroke in ATHENA: a placebo-controlled, double-blind, parallel-arm trial to assess the efficacy of dronedarone 400 mg BID for the prevention of cardiovascular hospitalization or death from any cause in patients with atrial fibrillation/atrial flutter. *Circulation*. 2009:29;120:1174–1180.
47. Terasawa T, Balk EM, Chung M, Garlitski AC, Alsheikh-Ali AA, Lau J, Ip S. Systematic review: comparative effectiveness of radiofrequency catheter ablation for atrial fibrillation. *Ann Intern Med*. 2009;151:191–202.
48. European Heart Rhythm Association (EHRA); European Cardiac Arrhythmia Society (ECAS); American College of Cardiology (ACC); American Heart Association (AHA); Society of Thoracic Surgeons (STS); Calkins H, Brugada J, Packer DL, Cappato R, Chen SA, Crijns HJ, Damiano RJ Jr, Davies DW, Haines DE, Haissaguerre M, Iesaka Y, Jackman W, Jais P, Kottkamp H, Kuck KH, Lindsay BD, Marchlinski FE, McCarthy PM, Mont JL, Morady F, Nademanee K, Natale A, Pappone C, Prystowsky E, Raviele A, Ruskin JN, Shemin RJ. HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for personnel, policy, procedures and follow-up: a report of the Heart Rhythm Society (HRS) Task Force on catheter and surgical ablation of atrial fibrillation [published correction appears in *Heart Rhythm*. 2009;6;148]. *Heart Rhythm*. 2007;4:816–861.
49. Catheter Ablation Versus Anti-Arrhythmic Drug Therapy for Atrial Fibrillation Trial (NCT00911508). ClinicalTrials.gov Web site. http://www.clinicaltrials.gov/. Accessed October 11, 2010.
50. Damiano RJ Jr, Schwartz FH, Bailey MS, Maniar HS, Munfakh NA, Moon MR, Schuessler RB. The Cox maze IV procedure: predictors of late recurrence. *Thorac Cardiovasc Surg*. 2011;141:113–121.
51. Weimar T, Bailey MS, Watanabe Y, Marin D, Maniar HS, Schuessler RB, Damiano RJ Jr. The Cox-maze IV procedure for lone atrial fibrillation: a single center experience in 100 consecutive patients. *J Interv Card Electrophysiol*. February 22, 2011. doi:10.1007/s10840-011-9547-3. http://www.springerlink.com/content/e761u08283750648. Accessed March 26, 2011.
52. Curtis LH, Whellan DJ, Hammill BG, Hernandez AF, Anstrom KJ, Shea AM, Schulman KA. Incidence and prevalence of heart failure in elderly persons, 1994–2003. *Arch Intern Med*. 2008;168:418–424.
53. Mehra R, Benjamin EJ, Shahar E, Gottlieb DJ, Nawabit R, Kirchner HL, Sahadevan J, Redline S; Sleep Health Study. Association of nocturnal arrhythmias with sleep-disordered breathing: the Sleep Heart Health Study. *Am J Respir Crit Care Med*. 2006;173:910–916.
54. Gami AS, Hodge DO, Herges RM, Olson EJ, Nykodym J, Kara T, Somers VK. Obstructive sleep apnea, obesity, and the risk of incident atrial fibrillation. *J Am Coll Cardiol*. 2007;49:565–571.
55. Kanagala R, Murali NS, Friedman PA, Ammash NM, Gersh BJ, Ballman KV, M. Shamsuzzaman AS, Somers VK. Obstructive sleep apnea and the recurrence of atrial fibrillation. *Circulation*. 2003;107:2589–2594.
56. Monahan K, Storfer-Isser A, Mehra R, Shahar E, Mittleman M, Rottman J, Punjabi N, Sanders M, Quan SF, Resnick H, Redline S. Triggering of nocturnal arrhythmias by sleep-disordered breathing events. *J Am Coll Cardiol*. 2009;54:1797–1804.
57. Stevenson IH, Roberts-Thomson KC, Kistler PM, Edwards GA, Spence S, Sanders P, Kalman JM. Atrial electrophysiology is altered by acute hypercapnia but not hypoxemia: implications for promotion of atrial fibrillation in pulmonary disease and sleep apnea. *Heart Rhythm*. 2010;7:1263–1270.
58. Sin DD, Fitzgerald F, Parker JD, Newton G, Floras JS, Bradley TD. Risk factors for central and obstructive sleep apnea in 450 men and women with congestive heart failure. *Am J Respir Crit Care Med*. 1999;160:1101–1106.
59. Gami AS, Pressman G, Caples SM, Kanagala R, Gard JJ, Davison DE, Malouf JF, Ammash NM, Friedman PA, Somers VK. Association of atrial fibrillation and obstructive sleep apnea. *Circulation*. 2004;110:364–367.
60. Estes NAM 3rd, Halperin JL, Calkins H, Ezekowitz MD, Gitman P, Go AS, McNamara RL, Messer JV, Ritchie JL, Romeo SJ, Waldo AL, Wyse DG. ACC/AHA/Physician Consortium 2008 clinical performance measures for adults with nonvalvular atrial fibrillation or atrial flutter: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures and the Physician Consortium for Performance Improvement (Writing Committee to Develop Clinical Performance Measures for Atrial Fibrillation): developed in collaboration with the Heart Rhythm Society. *Circulation*. 2008;117:1101–1120.

KEY WORDS: AHA Scientific Statements ■ atrial fibrillation ■ atrium ■ epidemiology ■ prevention ■ risk factors

Financial Analysis; Emotional Health; Measurement Issues

Workplace Telecommunications Technology to Identify Mental Health Disorders and Facilitate Self-Help or Professional Referrals

Ramesh Farzanfar, PhD; Steven E. Locke, MD; Timothy C. Heeren, PhD; Allison Stevens, MA; Louis Vachon, MD; Mai Khoa Thi Nguyen, BA; Robert H. Friedman, MD

Abstract

Purpose. *Test the feasibility and impact of an automated workplace mental health assessment and intervention.*

Design. *Efficacy was evaluated in a randomized control trial comparing employees who received screening and intervention with those who received only screening.*

Setting. *Workplace.*

Subjects. *463 volunteers from Boston Medical Center, Boston University, and EMC and other employed adults, among whom 164 were randomized to the intervention ($N = 87$) and control ($N = 77$) groups.*

Intervention. *The system administers a panel of telephonic assessment instruments followed by tailored information, education, and referrals.*

Measures. *The Work Limitation Questionnaire, the Medical Outcomes Questionnaire Short Form-12, the Patient Health Questionnaire-9, question 10 from the Patient Health Questionnaire to measure functional impairment, and the Perceived Stress Scale-4 and questions written by study psychiatrists to measure emotional distress and social support respectively. The WHO-Five Well-being Index was administered to measure overall well-being.*

Analysis. *Independent sample t-tests and χ^2 tests as well as mean change were used to compare the data.*

Results. *No significant differences on 16 of the 20 comparisons at 3- and 6-month time points. The intervention group showed a significant improvement in depression ($p \leq .05$) at 3 months and on two Work Limitation Questionnaire subscales, the Mental-Interpersonal Scale ($p \leq .05$) and the Time and Scheduling Scale ($p \leq .05$), at 3 and 6 months respectively with a suggestive improvement in mental health at 6 months ($p \leq .10$).*

Conclusions. *This is a potentially fruitful area for research with important implications for workplace behavioral interventions. (Am J Health Promot 2011;25[3]:207–216.)*

Key Words: Automated Mental Health Assessment, Workplace Mental Health Assessment, Computers in Mental Health Assessment, Behavioral Telehealth, Prevention Research. Manuscript format: research; Research purpose: intervention testing; Study design: randomized trial; Outcome measure: productivity, morbidity, behavioral; Setting: workplace; Health focus: stress management; Strategy: education; skill/building/behavior change; Target population: adults; Target population circumstances: employees

Ramesh Farzanfar, PhD, is an Assistant Professor of Medicine, Medical Information Systems Unit, Boston University School of Medicine, Boston Medical Center, Boston, Massachusetts. Steven E. Locke, MD, is a Consulting Psychiatrist, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts. Timothy C. Heeren, PhD, is a Professor of Public Health, Boston University School of Public Health, Boston, Massachusetts. Allison Stevens, MA, is a Research Technician, and Mai Khoa Thi Nguyen, BA, is a Research Assistant, Massachusetts General Hospital, Charlestown, Massachusetts. Louis Vachon, MD, is a Professor of Psychiatry, Boston University School of Medicine, Boston, Massachusetts. Robert H. Friedman, MD, is a Professor of Medicine and Public Health and Chief, Medical Information Systems Unit, Boston University and Boston Medical Center, Boston, Massachusetts.

Send reprint requests to Ramesh Farzanfar, PhD, Assistant Professor of Medicine, Medical Information Systems Unit, Boston University School of Medicine, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118; rfarzanf@bu.du.

This manuscript was submitted January 18, 2010; revisions were requested June 25, July 16, and July 28, 2010; the manuscript was accepted for publication August 9, 2010.

0890-1171/11/$5.00 + 0
DOI: 10.4278/ajhp.100118-QUAN-14

INTRODUCTION

Research shows that approximately 30% of American adults suffer annually from mental health disorders[1] and that these disorders are usually chronic or recurring.[2] Furthermore, almost half of the affected individuals have two or more such disorders.[1] More importantly, it has been estimated that only 50% of those afflicted with a serious mental health disorder seek help.[3] The collateral negative impact of undiagnosed, untreated, or poorly treated mental health disorders includes marital and family instability[2,4,5] as well as adverse impacts on the social and economic fabric of the local community and the society at large.[6] The staggering cost is economically important because most working individuals spend a significant amount of time in the workplace. In most circumstances, however, mental health issues are neglected in the workplace until they cause problems such as reduced productivity and/or absenteeism.[2,4,5] It is not known to what extent employers' initiatives such as establishment of Employee Assistance Programs or other venues such as annual depression screening days have addressed the substantially high barriers to the detection and treatment of mental health disorders in the workplace. There has also been a dearth of empirical evidence to demonstrate the feasibility of programs initiated by employers to help employees address mental and emotional health issues, with some exceptions. For example, the study carried out by Wang et al.[7] indicates that organized and integrated

interventions to screen and manage employees with depression show positive results. Wang et al.[7] evaluated the effects of a depression outreach treatment program on workplace outcomes. The intervention consisted of a telephonic outreach and care management program that motivated workers to enter outpatient treatment and at the same time monitored treatment while providing recommendations to the providers. The results demonstrated that workers in the intervention group had significantly lower depression symptoms ($p = .009$) and higher job retention ($p = .02$) as well as more hours worked ($p = .02$). The Wang et al.[7] study demonstrates that an organized approach to addressing mental and emotional health issues in the workplace is possible and recommended. People with symptoms of a mental health disorder are often reluctant to seek treatment because of the stigma attached to mental disorders and fear of the negative personal and social consequences of being ''branded'' with a mental health illness.[8] In the workplace, these barriers are compounded but might be mitigated if there were a confidential, easily accessible, and low-cost method to screen affected individuals and to provide an effective and efficient intervention. There is evidence that computers are considered by many users to be more trustworthy and less judgmental than humans, especially when their use involves inquiring about personal, sensitive, and/or uncomfortable topics.[9,10]

This paper describes the results of a clinical trial conducted with a computer telephony (interactive voice response [IVR]) system for detecting undiagnosed and/or untreated mental health disorders in the workplace and for helping those who screen positive for a given disorder to obtain effective treatment. Our objective was to reach such individuals directly in a convenient setting such as the workplace or at home rather than in the offices of a mental health provider. The system we developed, Telephone-Linked Communications for Detection of Mental Health Disorders (TLC-Detect), was initially deployed for use in workplace settings.

METHODS

Design

The entire study, including eligibility screening and outcome data collection, was carried out over the telephone at baseline and 3- and 6-month follow-ups. After eligibility screening, baseline data were collected from study participants, who were subsequently randomized and connected to the automated program to receive assessment for mental health disorders (all subjects) and intervention (only experimental subjects). The data derived from the control group subjects were used in a comparison to determine the efficacy of the TLC-Detect intervention. From a social perspective, the goal of the system was to improve workplace productivity by reducing the impact of mental health disorders on absenteeism and presenteeism (being at work but not being fully productive). As a result, the study's outcomes of interest were both work-related (absenteeism and presenteeism) and individual-based (physical and mental well-being).

Sample

The study protocol, including the sampling procedures described below, was approved by the Boston University Medical Campus' Institutional Review Board. Once system development was completed, the study was advertised and potential participants called our research headquarters to volunteer for participation. The study was advertised at Boston Medical Center and Boston University in addition to Florida Power & Light, EMC Corporation, and later, because of recruitment difficulties, to employed adults at large. We received the highest number of volunteers from Boston Medical Center and Boston University (253), followed by other companies (193), EMC Corporation (15), and Florida Power & Light (2).

To be eligible for entry into the trial, potential participants had to satisfy the following inclusion criteria: (1) ability to speak and understand conversational English, (2) 18 years of age or older, (3) access to a touch-tone telephone, (4) not undergoing mental health treatment or currently taking a medication prescribed for mental health treatment, and (5) experiencing some type of emotional distress as indicated by scoring positive on the WHO-5 Well-being Index[11–13] and the Functional Impairment question (item 10) from the Patient Health Questionnaire (PHQ) (''If you said *yes* to any of the five questions, how difficult have those problems made it for you to do your work, take care of things at home, or get along with other people''?). One hundred sixty-four participants were randomized in the clinical trial and received full assessment by TLC-Detect, with 87 and 77 randomized into the intervention and control groups respectively. Of these, 152 study participants (intervention = 77, control = 75) completed the study and 12 were lost to follow-up or dropped out of the study (8 before the 3-month and 4 before the 6-month outcome data collection). Of these 10 were in the intervention group and 2 were in the control group.

Measures

The WHO-5 Well-being Index was used to assess eligibility. WHO-5 was derived from a larger rating scale developed for a World Health Organization project on quality of life in patients suffering from diabetes.[16] The instrument has been successfully utilized in psychiatric care with good internal consistency (Cronbach $\alpha = .91$).[17] It has also shown a sensitivity and specificity of .93 and .64 respectively. WHO-5 covers mood (good spirits, relaxation), vitality (being active and waking up fresh and rested), and general interests (being interested in things) on a six-point scale from ''All of the time'' to ''At no time.''[14,15] The total WHO-5 score ranges from 0 to 25, with 25 being the most desired score.

We used several instruments to measure the impact of the intervention on study participants' work productivity (e.g., absenteeism and presenteeism) and physical and mental well-being as well as change in symptom severity (positive or negative) during the study. Outcomes were measured using the Work Limitation Questionnaire, a 29-item instrument with five different scales that address the user's productivity at work;[16] a four-item scale measures absenteeism, and the additional four scales (Mental Demands

scale, Output Demands scale, Physical Demands scale, and Time and Scheduling scale) focus on presenteeism, i.e., the user's actual self-reported impairment of productivity when at work. This instrument has good internal consistency, with four of the five scales (Time Demands, Physical Demands, Mental-Interpersonal Demands and Output Demands) having a Cronbach α of $\geq .90$.[19] The possible range for the total Work Limitation Questionnaire score is from 0 to 100, with a lower score being more desirable.

In addition, we used the Medical Outcomes Questionnaire Short Form-12 (SF-12) to measure physical and mental well-being. The SF-12 is a self-report measure of patient health status and covers a number of physical and health domains. The SF-12 includes all eight SF-36 scales (physical function, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health). Two items are used to measure four of the constructs and one item is used to measure the other four constructs. Research has shown that the SF-12 replicates 90% of the variance in both the physical component summary (PCS) and mental component summary (MCS) scores of the SF-36. The test-retest reliability of SF-12 is estimated at .89 for the PCS and .76 for the MCS.[20] Both the MCS and the PCS range from 0 to 100, with higher scores being more desirable. Both the Work Limitations Questionnaire and the SF-12 have been validated among groups with chronic conditions and mental health disorders.

Furthermore, we collected data on symptom severity, including administration of the PHQ-9[21] to measure each participant's depression level. The PHQ Total Depression score ranges from 0 to 27, where a higher score is worse. During 3- and 6-month outcome data collection, we also asked study participants whether the symptoms for each specific disorder had improved, remained the same, or worsened. We assessed functional impairment by asking question 10 from the PHQ combined with questions from the Perceived Stress Scale-4[22] to determine the degree to which employees felt their emotional distress interfered with their daily life. In addition, we inquired about social support, using three questions written by our study psychiatrists, to account for possible variations in the support study participants may have received from friends and family during the study. Also, we repeated the administration of the WHO-5 Well-being Index during the 3- and 6-month follow-up outcome data collection (as mentioned above the instrument was used to determine eligibility for study entry).

Finally, upon the completion of the study, we administered the Health Technology Questionnaire (HTQ) to explore the usability of the system. The HTQ has been developed by the first author and is designed to measure users' experience of automated, multi-contact health behavior interventions.[23] This instrument was developed based on two consecutive qualitative evaluations of two different computer telephony (Telephone-Linked Communications [TLC]) systems. The HTQ measures user satisfaction and perceptions of the system's impact on personal health behavior on a 5-point scale from ''Not at All'' to ''Very'' or from ''Strongly Agree'' to ''Strongly Disagree,'' with higher scores being desirable. For example, one of the subscales addresses satisfaction and includes questions such as ''How satisfied were you''? or ''Was the information useful''?[23] Another subscale addresses behavioral constructs, e.g., self-efficacy, motivation, and awareness. The instrument's subscales have displayed good to high reliability, with Cronbach α higher than .7 on 11 of the 12 different subscales.[23] In this paper, we report HTQ's percentages with favorable responses.

Intervention

The TLC-Detect system is based on an IVR technology called TLC.[11,12] TLC communicates with patients/users to promote health and prevent disease. TLC-Detect was designed as a mental health application of TLC to screen for undiagnosed and/or untreated mental health problems and help determine feasible self-management or professional care options.

We designed the system to be an automated mental health screening and counseling program that employees could access from any phone—home, work, etc. The participants used the telephone keypad to record their responses, which eliminated any confidentiality concerns that come from speaking into a telephone receiver. Either the user or the system could initiate the telephone call. At the beginning of each interaction, the user is asked to enter a unique password, which allows the system to easily recognize individual users. TLC-Detect used a prerecorded, digitized voice of a female voice actor. We coached the voice actor's tone and pitch to ensure that her delivery was appropriate for each specific disorder or condition.

The system was designed to include three modules: the screening module, the intervention module, and the intervention follow-up module. Both control and intervention group participants used the screening module; however, only participants in the intervention group were provided with the intervention module and the intervention follow-up module. These contained specific information, education, and referrals for self-help or professional assistance relevant to a specific disorder. In addition, the intervention group participants received monthly follow-up calls to monitor their progress and adherence to the system's advice. Because of ethical considerations, participants in the control arm of the study who reported symptoms during assessment were briefly advised by the system to confer with their clinicians about their symptoms. However, no information, education, or referrals were provided to the control group participants. Upon completion of the study, the study staff reminded subjects in the control arm of the study to visit their physicians and discuss their emotional distress.

The Screening Module

The screening module assesses users for mental health disorders that are known to reduce employee productivity, including major and minor depression (as well as postpartum depression and depression due to acute bereavement), dysthymia (a milder but chronic form of depression), bipolar disorder, generalized anxiety disorder, somatization, posttraumatic stress disorder, social phobia, panic disorder,

acute stress disorder, suicidal ideation, violence, alcohol and drug problems (including abuse and dependence), and general stress such as that related to family, marital, work, or school as well as financial problems.[24]

This module uses instruments that have been used in valid and reliable ways in other studies. The most important of these instruments is the PHQ, which is the self-administered version of the Primary Care Evaluation of Mental Disorders.[21] Additional instruments used by the system include the Impact of Events Scale,[25] the Acute Stress Disorder Scale,[26] the Social Phobia Inventory,[27] and the Mood Disorder Questionnaire.[28]

Although the assessment module screens users for mental health symptoms, it is not intended to be a diagnostic program and should not be considered a substitute for a clinician diagnosis. The ultimate goal of the program is to motivate the user to seek help. The system compares patterns of users' self-reported symptoms with those symptom patterns observed in reference populations of individuals known to be affected by a specific disorder, i.e., diagnosed by a mental health clinician. The system detects symptoms characteristic to a specific disorder, and identifies the severity level of the disorder based on the scaled symptom severity assessment. This includes detecting subsyndromal levels for each identified disorder, i.e., when a person's symptoms do not meet criteria for a disorder but the score is severe enough to cause problems. Symptom severity cutoffs that define a syndrome level are based on the recommendations made by the developers of each screening instrument. In addition, the system measures an employee's level of social support, the employee's daily life stressors, and his/her functional impairment due to mental health problems,[24] particularly because these factors have been shown to have a considerable impact on the severity of a mental health disorder.[29] These factors also play a role in determining the severity level of an employee's distress and therefore affect the content of the intervention.

Subsequently, the system provides appropriate information and advice for any positive screenings of a disorder and its corresponding severity level. First, the user is advised that his/her symptoms are similar to those seen in patients with that disorder. Next, the recommendations are tailored to the symptom severity level. For example, to a person with a symptom pattern consistent with those of severe PTSD, the system will say, ''It is *critical* that you adhere to the following recommendations,'' and will then enumerate what the person needs to do.[24] In contrast, to a person with symptoms implying subsyndromal depression, the system might say, ''I *suggest* that you follow my advice with regards to the following …''

The screening module was designed to function in a hierarchical manner. The first level of screening includes questions for all users. The instruments that are used at this level are mainly derived from the PHQ, minus a few questions that were deemed inappropriate for the study population. Employees with positive responses to this level of screening proceed to the second level of more disorder-specific and in-depth screening by additional screening instruments. For example, an employee who screens positive for depression at the first level will be asked questions that distinguish different types of depression and help identify whether suicidality is present. If suicidality is present, a third tier of screening assesses whether the employee is experiencing suicidal ideation, i.e., whether the employee is only thinking about suicide or has engaged in suicide planning, i.e., has thought about a plan.[24] If a person has been planning a suicide, appropriate action is taken both by the system, which encourages the person to immediately go to the nearest emergency room, and the study staff, who will notify the house psychiatrist.

Because the system is designed to take into account severity, i.e., intensity and/or frequency of symptoms, most of the disorders are characterized as mild, moderate, or severe. However, when alcohol and/or drug problems are identified, the system assigns severity levels unique to substance abuse: risky alcohol/drug use, alcohol/drug abuse, and alcohol/drug dependence. These classifications enable the system to provide appropriate intervention for each disorder category.[24] The system also screens employees to determine whether they have been victims of violence. Because of the sensitivity of this question, reluctant users are allowed to skip answering this question. The system keeps track of those who skip this question in order to take this into consideration for future modifications of the TLC-Detect system.[24]

Another important function of the assessment module is the identification of up to eight different, coexisting disorders, i.e., occurrence of more than one disorder per employee/user. In addition, to address the challenge of prioritizing intervention delivery, the system ranks coexisting disorders based on their relative severity or seriousness.[24] For example, in our model, bipolar disorder is deemed to be a more severe or serious disorder than dysthymia. Similarly, if an employee reports being the victim of violence as well as having symptoms consistent with major depression and social phobia, the system will rank the disorders based on relative severity and adjusts the intervention materials accordingly. As a result, because violence is defined as an immediate concern, it takes clinical precedence and the system will first provide information and advice about violence. Once the employee has listened to the intervention for violence and is ready to hear more, he or she will then hear about major depression, because major depression carries more risk compared to social phobia, and lastly, the system will give intervention information about social phobia.

The screening module also contains a submodule for ''Unspecified Emotional Distress''[24] to address situations in which a person has a high level of life stressors or significant functional impairment, but does not have a positive score for any of the listed mental health disorders. This submodule contains an intervention for general distress. It was particularly important to offer some type of intervention to users who were not positive for any specific disorder but during eligibility screening reported having a specific threshold level of emotional distress. Depending on the number of comorbid disorders, a screening session with

the system takes between 30 and 90 minutes and can be divided into two different sessions.

The Intervention Module

This module provides information, education, and referral information to experimental subjects. The intervention is offered to subjects immediately after they complete the assessment module. Specifically, the two important components of the intervention module are the education and referral submodules. If the user has a criterion-level symptom severity pattern for any disorder, the system provides extensive information about that disorder, including its symptoms, natural history, and available treatments. The referral submodule contains disorder-specific information on both self-management and professional help appropriate to the level of its severity as determined by the system's assessment. Employee/users listen to information about the treatment options, treatment advantages and limitations, and advice on how to access self-help resources. Self-management resources include self-help workbooks, learning and using stress management techniques, participating in support groups, beginning an exercise regimen and/or adopting a healthier diet, etc. Further information on each subtopic is provided. For example, the section on stress management techniques includes instructions on mediation, progressive muscle relaxation, autogenic training, and breathing exercises. When the screening assessment warrants it, users are also given information on how to find appropriate professional treatment. The referral module provides information about available professional help (clinical psychologists, psychiatrists, or social workers) and targets specific geographic areas, employers, and health plans to guide users to find a mental health professional. However, the system does not provide specific clinicians' names and telephone numbers. The referral module also includes individual and group therapy options, based on the screening assessment of disorder severity. For example, to an employee deemed likely to have moderate major depression, the system recommends self-management options such as exercise, healthy diet, meditation, an appropriate workbook, and a support group as well as professional assistance such as group therapy and/or Employee Assistance Program for evaluation.

To ensure that referrals are properly assigned and target each disorder appropriately, the system designers created the Treatment Intensity Adjuster (TIA),[29] which measures the impact of each disorder on a person's personal and social life by assessing his/her functional impairment, daily stressors, and social support level.[24] A high TIA score increases the assigned severity level of each disorder by one level.[29] This enables the system's intervention module to recommend appropriate treatment strategies to each person based on the severity (seriousness) level of his/her disorder. Ideally, the assessment and intervention are provided to users in one 30- to 90-minute session. The average amount of time that a potential user spends on the screening and intervention modules is about 60 minutes, but it varies depending on the number and severity of the identified disorders. If a person has several disorders and thus needs to listen to more information, the session may be divided to two in order to reduce the time burden on the user.

The Intervention Follow-up Module

This module may track employee users on a monthly basis for any given period of time. The principal objective of the intervention follow-up calls is to ensure that users have adhered to the system's advice and sought professional assistance or engaged in self-help. For users who did not seek professional treatment or engage in self-help, the intervention follow-up module provides tailored educational material, which might include description of the disorder and its treatment options. In addition to checking whether a user is following the system's recommendations, the follow-up module is also used to provide additional information to a user about his/her condition(s). Similar to the intervention module, the users again have the choice to spread out the information in the intervention follow-up module into multiple sessions to reduce the time burden. If the users have already listened to *all* of the information offered by the system, they have the option to listen to any information again. The system offers a variety of optional modules, so users are unlikely to complete all the modules in the system. For example, there is an optional book recommendation module that lists descriptions and titles of informative books about any disorder that is of concern to the user. In addition to the above functions, the intervention follow-up module also reviews recommendations and barriers to adherence with nonadherent individuals. Once a barrier is discussed, the system offers appropriate advice to enable these employees to overcome problems or difficulties that prevented their adherence to the system's recommendations. The intention is to empower and motivate users to become actively involved in the management of their mental health. A final function of the intervention follow-up module is to determine the accuracy of the system's assessment by asking whether users have discussed the identified disorders with their health care providers. If so, the system then asks whether the clinicians agreed or disagreed with the assessment.

Analysis

We compared changes in productivity (defined by Work Limitation Questionnaire scores) and changes in mental health symptomatology (SF-12, PHQ, and Stress Level Questionnaire) from baseline to 3 months (initial effects) and baseline to 6 months (longer-term effects), using separate independent sample *t*-tests.

RESULTS

In order to be eligible for the study, all subjects had to have some level of mental/emotional distress at the time of their enrollment. This was determined by the administration of the WHO-5 Well-being Index. Subsequently, all participants were screened by the system and then randomized. This means that the majority of the study participants were identified with a specific disorder or comorbid disorders with the exception of 6% (intervention) and 5% (control) who were identified with unspecified emotional distress. However, after screening by the system only participants random-

Table 1
Description of Study Sample at Baseline*

	Control (N = 78)	Intervention (TLC Group) (N = 89)	Significance†
Current age, y, mean (SD)	39.2 (11.5)	39.0 (10.4)	0.9315
Gender, No. (%)			
Male	17 (21.8)	24 (27.0)	
Female	61 (78.2)	65 (73.0)	0.4385
Married/living with partner, No. (%)			
Yes	43 (55.1)	42 (47.2)	
No	35 (44.9)	47 (52.8)	0.3060
Education, No. (%)			
College graduate	49 (62.8)	54 (60.7)	
Less than college	29 (37.2)	35 (39.3)	0.7759
Race, No. (%)			
White	45 (57.7)	49 (55.1)	
Black or African-American	23 (29.5)	30 (33.7)	0.8304
Other	10 (12.8)	10 (11.2)	
Hispanic or Latino, No. (%)			
Yes	11 (14.1)	5 (5.6)	
No	67 (85.9)	84 (94.4)	0.0631
Annual salary, No. (%)			
≤$49,999	53 (67.9)	59 (67.8)	
$55,000+	25 (32.1)	28 (32.2)	0.999

* TLC indicates Telephone-Linked Communications.
† Comparison of intervention vs. control groups via independent sample *t*-test for measurement variables and the χ^2 test for categorical variables.

ized to the experimental arm received the intervention.

Complete data were collected at baseline, 3 months, and 6 months for 91% of the randomized subjects (total: 152 out of 164; control group: 75 out of 77; intervention group: 77 out of 87). Baseline characteristics of the study sample are presented in Table 1. There were no significant demographic differences between the two study groups at baseline. In both groups, subjects tended to be in their 30s and 40s, with about 75% female and 60% college educated, and with minority representation.

The study lost 12 participants to follow-up. These participants did not substantially differ from those followed on demographic or mental health characteristics. Of these, one participant in the control group did not complete both outcome data collections, compared with seven intervention group participants. In addition, one participant in the control group did not complete the 6-month outcome data collection, compared to three in the intervention group. The higher number of dropouts among the intervention group participants may be an indication of the greater time commitment that was expected from subjects in the intervention arm of the study.

All participants in the intervention group received a brief follow-up call once a month for a total of 6 months to verify that they followed at least one of the program's recommendations and determine whether they have sought help suggested by the system. Of the 87 participants in the intervention group, 68 (78%) responded to at least one follow-up call, with 44 (51%) responding to five or six of the six scheduled calls.

Comparison of the two study groups' baseline mental health and work-related characteristics is presented in Table 2. There were no significant differences between the study groups at baseline. Given the comparability of study groups on demographic, mental health, and work-related characteristics at baseline, these factors should not confound our outcome comparison of the study groups; thus, our primary analyses focused on comparison of the two study groups without adjusting for differences in baseline characteristics.

Changes in outcome measures at 3 and 6 months are presented in Table 3. We examined changes in five work-related outcomes and five health/mental health–related outcomes at two time points (20 comparisons in all). There were no significant differences between the intervention and control groups on 16 of these comparisons, and there was a significant advantage in the intervention group on three of these comparisons with a suggestive result on a fourth comparison.

Those in the intervention group showed a significantly greater reduction in depression ($p \leq .05$) at 3 months, a suggestive improvement in general mental health at 6 months ($p \leq .10$), as well as a significantly greater improvement on two Work Limitation Questionnaire subscales, the Mental-Interpersonal Scale ($p \leq .05$) at 3 months and the Time and Scheduling Scale ($p \leq .05$) at 6 months.

Table 2
Productivity and Mental Health Well-Being Values at Baseline*

	Control (N = 78)	Intervention (TLC Group) (N = 89)	Significance†
Work Limitation Questionnaire, mean (SD)‡			
Time and Scheduling Scale score	38.2 (19.3)	39.6 (20.2)	0.6438
Physical Scale score	20.4 (18.8)	23.7 (21.3)	0.3026
Mental-Interpersonal Scale score	37.2 (18.3)	36.2 (18.3)	0.7446
Output Scale Score	35.2 (23.5)	34.3 (21.2)	0.7895
Productivity Index	10.1 (5.0)	9.9 (4.8)	0.8012
SF-12§			
General Health, No. (%)			
Excellent	5 (6.4)	6 (6.7)	0.999
Very good	32 (41.0)	35 (39.3)	
Good	32 (41.0)	34 (38.2)	
Fair	7 (9.0)	11 (12.4)	
Poor	2 (2.6)	3 (3.4)	
Physical Health Scale, mean (SD)	49.5 (10.1)	47.7 (11.1)	0.2866
Mental Health Scale, mean (SD)	36.9 (10.9)	37.5 (9.3)	0.7276
PHQ-9‖			
Total Depression score, mean (SD)	7.7 (4.9)	7.9 (5.3)	0.7398
Depression severity level, No. (%)			0.2161
Severe depression	2 (3.9)	4 (4.6)	
Moderate depression	11 (14.3)	12 (13.8)	
Minor depression	4 (5.2)	7 (8.1)	
Subsyndromal depression	20 (26.0)	10 (11.5)	
No depression	39 (50.7)	54 (62.1)	
Stress Questionnaire level, No. (%)			
High stress (11+)	7 (9.1)	11 (12.6)	0.4199
Moderate stress (6–10)	46 (59.7)	41 (47.1)	
Low stress (2–5)	21 (27.3)	29 (33.3)	
No stress (0–1)	3 (3.9)	6 (6.9)	
Stress Questionnaire score, mean (SD)¶	2.3 (0.7)	2.3 (0.8)	0.4628
WHO-5 Well-being Index#			
Total score, mean (SD)	10.3 (4.6)	10.4 (4.1)	0.9091

* TLC indicates Telephone-Linked Communications; SF-12, Medical Outcomes Questionnaire Short Form-12; and PHQ-9, Patient Health Questionnaire-9.
† Comparison of intervention vs. control via independent sample *t*-test for measurement variables and χ^2 test for categorical variables.
‡ Work Limitation Questionnaire scales from 0 to 100 with higher values indicating greater problems.
§ SF-12 scores from 0 to 100 with higher values indicating better functioning.
‖ PHQ scale from 0 to 27 with higher values indicating more depression.
¶ Stress Questionnaire score from 1 to 16 with higher values indicating greater stress.
# WHO-5 score from 0 to 25 with higher values indicating better functioning.

We also examined system usability by administering a structured questionnaire (the HTQ)[21] to all study participants upon the completion of the study. The HTQ measures overall opinion of users in addition to using a number of behavioral constructs. Based on our analysis, intervention group subjects found the system easy to use (84% reporting very easy or somewhat easy to use), friendly (80% very or somewhat friendly), appropriately paced (67% reporting the pace was just right), and informative (76% reporting very or somewhat informative). Furthermore, 65% reported that the system was very or somewhat useful and 47% agreed that the system reduced their visit time with their doctor.[21]

DISCUSSION

The results of this randomized clinical trial indicate that an automated telephony system might be helpful in bringing about small improvements in clinical and work-related outcomes for individuals with common mental health disorders. Our intervention showed a significant improvement at either 3 or 6 months on at least three (and a suggestive fourth) of 10 outcome measures examined including both work-related (Work Limitation Questionnaire Time and Scheduling Scale score) and mental health–related (SF-12 Mental Health Scale, PHQ Total Depression scale) outcomes. Finally, based on the results of our HTQ, the system was usable and acceptable to most study participants, who did not

Table 3
Changes in Productivity and Mental Health Well-Being Values at 3 and 6 Months†

	Control (n = 75)		Intervention (n = 77)	
	3 mo	6 mo	3 mo	6 mo
Work Limitation Questionnaire‡				
Time and Scheduling Scale score	−9.9 (18.8)	−8.6 (21.4)	−15.3 (22.5)	−17.2 (23.6)*
Physical Scale score	−4.9 (16.7)	−4.2 (14.9)	−5.8 (26.8)	−7.3 (23.9)
Mental-Interpersonal Scale score	−8.3 (16.9)	−10.7 (21.6)	−14.6 (20.9)*	−15.0 (22.9)
Output Scale score	−8.2 (19.0)	−8.2 (21.5)	−11.7 (22.3)	−14.7 (27.1)
Productivity Index	−2.4 (4.1)	−2.7 (4.7)	−3.6 (4.9)	−4.1 (5.7)
SF-12§				
Physical Health Scale	+0.3 (7.9)	+1.2 (8.5)	+0.9 (9.9)	+2.1 (10.5)
Mental Health Scale	+5.1 (11.2)	+6.0 (12.7)	+7.7 (10.4)	+10.9 (10.1)**
PHQ-9‖				
Total Depression score	−0.1 (4.6)	−1.8 (4.5)	−1.5 (3.9)*	−2.2 (4.7)
Stress Questionnaire level				
Stress Questionnaire score¶	−1.0 (2.8)	−1.8 (3.1)	−1.5 (3.3)	−2.1 (3.4)
WHO-5 Well-being Index total score#	2.1 (5.4)	3.5 (7.1)	3.6 (6.1)	3.7 (6.8)

† Data tabled are mean (SD) change scores. Significant differences between intervention and control are indicated in bold. Comparison of intervention vs. control via independent sample *t*-test for measurement. SF-12 indicates Medical Outcomes Questionnaire Short Form-12; PHQ-9, Patient Health Questionnaire-9.
‡ Work Limitation Questionnaire scales from 0 to 100 with higher values indicating greater problems.
§ SF-12 scores from 0 to 100 with higher values indicating better functioning.
‖ PHQ scale from 0 to 27 with higher values indicating more depression.
¶ Stress Questionnaire score from 1 to 16 with higher values indicating greater stress.
# WHO-5 score from 0 to 25 with higher values indicating better functioning.
* $p < 0.05$, comparing mean change in TLC group vs. control.
** $p < 0.10$, comparing mean change in TLC group vs. control.

hesitate to provide it with sensitive and sometimes potentially stigmatizing information. Our findings, taken together with other work in this field,[30] indicate that this is a potentially fruitful area for research with important implications for behavioral interventions targeting work performance impaired by behavioral health disorders.

A myriad of IVR systems have been developed that are used in the provision of mental health services, such as automated interviews to assess depression[31] and anxiety,[32] prediction of suicide risk,[33] personality testing,[34] and others. Most of this research has focused in the area of automating standardized mental health questionnaires and testing them for reliability/validity.[31–34] In addition to mental health assessment, the technology has also been used to provide behavior therapy to patients with various affective disorders.[35,36] Many of these IVR studies, however, are not randomly controlled clinical trials. In fact, the few clinical trials that have used mental health self-assessment by computers (not IVR) have shown some benefits to users. For example, a randomized controlled evaluation of computerized assessment of common mental health disorders in primary care showed small improvements in clinical outcomes for patients with these disorders.[37] The similar results produced by this study shows that computers are of certain value in the applications for assessment and treatment of mental health disorders. Furthermore, a review monograph by Isaac Marks et al.[38] about computer-aided psychotherapy explored advantages and disadvantages of more than 175 different programs developed by researchers and clinicians around the world. They found out that some of these programs are more effective than others for a variety of reasons, including individual preferences (for a therapist or a computer), system design, method of use, and reason for use, among others.[38] This review is indicative of a trend but one that needs close evaluation to make these systems appropriate for widespread utilization by the public.

An important limitation of our study was recruitment problems, which led to the study's being underpowered to detect behavior change in a randomized intervention trial. With our sample size, moderately large effects, corresponding to Cohen's effect size[39] (difference in means divided by the common standard deviation) of d = .46, would be needed for 80% power capable of showing an intervention effect. Our study may have failed to detect a more modest intervention effect. Despite being underpowered and failing to detect more significant intervention effects, the interest given to the potential low cost and widespread access made possible by this type of easily scalable system is noteworthy. Another limitation with the potential to affect the outcome was the

fact that after the assessment was completed by the program, those in the control arm of the study were told by the automated system, ''The interview indicated that you may be suffering from some type of emotional distress. I recommend that you seek mental health care.'' This advice, requested by our Institutional Review Board, had to be provided to control group participants and was ethically required. Thus, those in the control group may have undertaken self-care efforts (e.g., behavioral activation or exercise or seeking care elsewhere) because of the effect of this potentially confounding yet ethical intervention.

SO WHAT? Implications for Health Promotion Practitioners and Researchers

What is already known on this topic?

Research has shown that automated behavioral interventions can yield clinical benefits. Studies targeting depression in primary care indicate that the principal beneficiaries of depression management programs are the employers. However, studies in the workplace that measure the impact of mental health disorders on productivity that use comprehensive automated behavioral health interventions are lacking.

What does this article add?

Our research showed that an automated behavioral health intervention can produce modest improvements in clinical and work-related outcomes. The article invites future work on exploration of human-computer interaction in mental health care with implications for augmenting traditional outpatient mental health care delivery. This innovative study calls for automated, scalable approaches that can reach a larger population of those in need of mental health services.

What are the implications for health promotion practice or research?

A practical contribution of our program is demonstrating the importance of targeting comorbid disorders and exploring whether improvement in one disorder increases the likelihood of achieving improvement in others.

We believe that our future research efforts should focus on a controlled trial of the system in which subjects will be randomized into two groups of automated system and human clinician. We hope to use a larger sample size in order to better detect differences between the intervention and control groups of each disorder. This will not only provide data on feasibility and impact but also measure the fidelity of the automated standardized instruments. Furthermore, because the program was uniquely designed to screen for a variety of common mental health disorders and to provide information and advice, future research should consider evaluation of the system with the general public and not just an employee population.

Acknowledgment

The development of the screening and intervention system was supported by a grant (Automated Assessment of Mental Health Disorders in the Workplace—R01 DP000116) from the Centers for Disease Control and Prevention.

References

1. Kessler RC, Kessler WT, Chiu O, et al. Prevalence, severity, and comorbidity of twelve-month DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R). *Arch Gen Psychiatry.* 2005;62:617–627.
2. Harnois G, Gabriel P. *Mental Health at Work: Impact, Issues and Good Practices.* Geneva, Switzerland: World Health Organization; 2000.
3. US Department of Health & Human Services. *Mental Health: A Report of the Surgeon General.* Rockville, Md: Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health; 1999.
4. Kessler RC, Berglund PA, Foster CL. The social consequences of psychiatric disorders. *Am J Psychiatry.* 1997;154:1405–1411.
5. Kessler RC, Heeringa S, Lakoma MD, et al. Individual and societal effect of mental disorders on earnings in the United States: results from the national comorbidity survey replication. *Am J Psychiatry.* 2008;165:663–665.
6. Robinson G, Chimento L, Bush S, Papy J. Comprehensive mental health insurance benefits: case studies. National Advisory Mental Health Council, Health Care Reform for Americans with Severe Mental Illness: report of the National Advisory Mental Health Council. Rockville, Md: Center for Mental Health Services, Substance Abuse and Mental Health Services Administrations; 2001.
7. Wang PS, Simon GE, Avorn J, et al. Telephone screening, outreach and care management for depressed workers and impact on clinical and work productivity outcomes: a randomized controlled trial. *JAMA.* 2007;298:1401–1411.
8. Corrigan P. How stigma interferes with mental health care. *Am Psychol.* 2004;59:614–625.
9. Wallace P. *The Psychology of the Internet.* Cambridge, United Kingdom: Cambridge University Press; 1999.
10. Locke SE, Kowaloff HB, Hoff RG, et al. Computer based interviews for screening blood donors for risk of HIV transmission. *JAMA.* 1993;268:1301–1305.
11. World Health Organization Regional Office for Europe and the International Diabetes Federation, Europe. Diabetes mellitus in Europe: a problem at all ages and in all countries: a model for prevention and self care. *G Ital Diabetol.* 1990;10(suppl):1–14. Available at: http://whqlibdoc.who.int/euro/-1993/EUR_ICP_NCD_023(1).pdf. Accessed December 12, 2008.
12. Newnham EA, Hooke GR, Page AC. Monitoring treatment response and outcomes using the World Health Organization's Wellbeing Index in psychiatric care. *J Affect Disord.* 2010;122:133–138.
13. Bech P. *Rating Scales for Psychopathology, Health Status and Quality of Life: A Compendium on Documentation in Accordance With the DSM-III-R and WHO Systems.* Berlin, Germany: Springer; 1993.
14. Bech P, Bech C, Gudex K, Johansen S. The WHO (Ten) Well-Being Index: validation in diabetes. *Psychother Psychosom.* 1996;65:183–190.
15. Bech P. *Quality of Life in the Psychiatric Patient.* London, United Kingdom: Mosby-Wolfe; 1998.
16. Lerner D, Amick BC, Rogers WH, et al. The Work Limitations Questionnaire. *Med Care.* 2001;39:72–85.
17. Lerner D, Reed J, Massarotti E, et al. The Work Limitation Questionnaire's validity and reliability among patients with osteoarthritis (OA). *J Clin Epidemiol.* 2002;55:197–208.
18. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1966;34:220–233.
19. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-ME: the PHQ primary care study. *JAMA.* 1999;282:1737–1744.
20. Cohen S, Williamson GM. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, eds. *The Social Psychology Of Health.* Newbury Park, Calif: Sage; 1988.
21. Stevens A, Migneault J, Fava J, et al. Health Technology Questionnaire: psychometric properties of a health behavior change measure [abstract]. *Ann Behav Med.* 2006;31(suppl):S126.
22. Friedman RH, Stollerman JE, Mahoney DM, Rozenblyum L. The virtual visit: using telecommunications technology to take care of patients. *J Am Med Inform Assoc.* 1997;4:413–425.

23. Friedman RH, Rozenblyum L, Belfer AS, et al. A telecommunications system to manage patients with chronic disease. *Medinfo: AMIA.* 1998;9:1330–1334.
24. Farzanfar R, Stevens A, Vachon L, et al. Design and development of a mental health assessment and intervention system. *J Med Syst.* 2007;31:49–62.
25. Horowitz M, Wilner N, Alvarez W. Impact of Event Scale: a measure of subjective stress. *Psychosom Med.* 1979;41:209–218.
26. Bryant RA, Moulds ML, Guthrie RM. Acute Stress Disorder Scale: a self-report measure of acute stress disorder. *Psychol Assess.* 2000;12:61–67.
27. Connor KM, Davidson JRT, Churchill LE, et al. Psychometric properties of the Social Phobia Inventory (SPIN). *Br J Psychiatry.* 2000;176:379–386.
28. Robert MA, Hirschfeld J, Williams BW, et al. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. *Am J Psychiatry.* 2000;157:1873–1875.
29. Abrams HS. Harvard Community Health Plan's Mental Health Redesign project, A managerial and clinical partnership. *Psychiatr Q.* 1993;64:13–31.
30. Cartreine J, Ahern DK, Locke SE. A road map to computer-based psychotherapy in the U.S. *Harv Rev Psychiatry.* 2010;18:89–95.
31. Mundt JC, Kobak KA, Taylor LV, et al. Administration of the Hamilton Depression Rating Scale using interactive voice response technology. *MD Comput.* 1998;15:31–39.
32. Kobak KA, Taylor LH, Dottl SL, et al. A computer-administered telephone interview to identify mental disorders. *JAMA.* 1997;278:905–910.
33. Erdman HP, Greist JH, Gustafson DH, et al. Suicide risk prediction by computer interview: a prospective study. *J Clin Psychiatry.* 1987;48:464–467.
34. Ben-Porath YS, Butcher JN. Computers in personality assessment: a brief past, an ebullient present, and an expanding future. *Comput Hum Behav.* 1986;2:167–182.
35. Baer L, Greist H. An interactive computer-administered self-assessment and self-help program for behavior therapy. *J Clin Psychiatry.* 1997;58(suppl):23–28.
36. Bachofen M, Nakagawa A, Marks IM, et al. Home self-assessment and self-treatment of obsessive-compulsive disorder using a manual and a computer-conducted telephone interview: replication of a UK-US study. *J Clin Psychiatry.* 1999;60:545–549.
37. Sharp LG, Bartholomew J, Pelosi AJ. Computerized assessment of common mental disorders in primary care: effects on clinical outcome. *Fam Pract.* 1996;13:120–126.
38. Marks I, Cavanagh K, Gega L. *Hands-on Help: Computer-Aided Psychotherapy.* New York, NY: Psychology Press; 2007.
39. Cohen J. *Stratified Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

CLINICAL TRIALS *DESIGN* *Clinical Trials* 2011; **8:** 183–195

A point-of-care clinical trial comparing insulin administered using a sliding scale *versus* a weight-based regimen

Louis D Fiore[a,b,c], Mary Brophy[a,c], Ryan E Ferguson[a,b], Leonard D'Avolio[a,d], John A Hermos[a,c,e], Robert A Lew[a,f], Gheorghe Doros[a,f], Chester H Conrad[a,c], Joseph A (''Gus'') O'Neil Jr[a], Thomas P Sabin[a], James Kaufman[a,c], Stephen L Swartz[a], Elizabeth Lawler[a,g,h], Matthew H Liang[a,i], J Michael Gaziano[a,g,h] and Philip W Lavori[j,k]

Background Clinical trials are widely considered the gold standard in comparative effectiveness research (CER) but the high cost and complexity of traditional trials and concerns about generalizability to broad patient populations and general clinical practice limit their appeal. Unsuccessful implementation of CER results limits the value of even the highest quality trials. Planning for a trial comparing two standard strategies of insulin administration for hospitalized patients led us to develop a new method for a clinical trial designed to be embedded directly into the clinical care setting thereby lowering the cost, increasing the pragmatic nature of the overall trial, strengthening implementation, and creating an integrated environment of research-based care.

Purpose We describe a novel randomized clinical trial that uses the informatics and statistics infrastructure of the Veterans Affairs Healthcare System (VA) to illustrate one key component (called the point-of-care clinical trial – POC-CT) of a 'learning healthcare system,' and settles a clinical question of interest to the VA.

Methods This study is an open-label, randomized trial comparing sliding scale regular insulin to a weight-based regimen for control of hyperglycemia, using the primary outcome length of stay, in non-ICU inpatients within the northeast region of the VA. All non-ICU patients who require in-hospital insulin therapy are eligible for the trial, and the VA's automated systems will be used to assess eligibility and present the possibility of randomization to the clinician at the point of care. Clinicians will indicate their approval for informed consent to be obtained by study staff. Adaptive randomization will assign up to 3000 patients, preferentially to the currently 'winning' strategy, and all care will proceed according to usual practices. Based on a Bayesian stopping rule, the study has acceptable frequentist operating characteristics (Type I error 6%, power 86%) against a 12% reduction of median length of stay from 5 to 4.4 days. The adaptive stopping rule promotes implementation of a successful treatment strategy.

[a]Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Cooperative Studies Program, VA Boston Healthcare System, Boston, MA, USA, [b]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA, [c]Department of Medicine, Boston University School of Medicine, Boston, MA, USA, [d]Center for Surgery and Public Health at Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, [e]Section of General Internal Medicine, Boston Medical Center, Boston, MA, USA, [f]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA, [g]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA, [h]Department of Medicine, Harvard Medical School, Boston, MA, USA, [i]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, MA, USA, [j]Health Research and Policy, Stanford University, Palo Alto, CA, USA, [k]Department of Statistics, Stanford University, Palo Alto, CA, USA

Author for correspondence: Louis Fiore, VA Healthcare System, 150 South Huntington Ave., Boston, MA 02130, USA
E-mail: louis.fiore@va.gov

10.1177/1740774511398368

Limitations Despite clinical equipoise, individual healthcare providers may have strong treatment preferences that jeopardize the success and implementation of the trial design, leading to low rates of randomization. Unblinded treatment assignment may bias results. In addition, generalization of clinical results to other healthcare systems may be limited by differences in patient population. Generalizability of the POC-CT method depends on the level of informatics and statistics infrastructure available to a healthcare system.

Conclusions The methods proposed will demonstrate outcome-based evaluation of control of hyperglycemia in hospitalized veterans. By institutionalizing a process of statistically sound and efficient learning, and by integrating that learning with automatic implementation of best practice, the participating VA Healthcare Systems will accelerate improvements in the effectiveness of care. *Clinical Trials* 2011; **8**: 183–195. http://ctj.sagepub.com

Introduction

Medical decision making is informed by clinical trials and observational studies. Randomization in clinical trials reduces or eliminates biases of observational studies, such as selection by indication and confounding from unmeasured prognostic factors that affect treatment decisions and outcomes. By their purpose, randomized clinical trials (RCTs) can be designed on a spectrum ranging from *pragmatic* (comparing effectiveness of interventions in the most realistic of situations and with diverse subjects) to *explanatory* (comparing efficacy in precisely described clinical situations and selected patients) [1,2]. The goal of explanatory trials is to better understand how and why an intervention works while pragmatic clinical trials are designed to provide information needed to assist healthcare providers make informed clinical decisions [3].

The *Pragmatic–Explanatory Continuum Indicator Summary (PRECIS)* is a measure of where on this continuum an individual trial is situated [4]. It takes under consideration the attributes of an RCT such as flexibility of the interventions, practitioner expertise required, eligibility criteria, intensity of follow-up and adherence monitoring, and the nature and scope of the primary outcome. RCTs are considered on the pragmatic end of the spectrum when these attributes are chosen to allow the trial to more closely mimic conditions encountered in the clinical care arena. Examples include eligibility criteria that reflect the patient population likely to receive the intervention, study investigators with expertise and experiences similar to the healthcare providers who will ultimately administer the treatments, treatment protocols that allow the flexibility required in routine clinical care, and outcome measures, and follow-up procedures that would be part of routine clinical care. Despite their reflection of routine clinical care, pragmatic trials are currently still complicated and expensive to implement, because of the use of dedicated study personnel to recruit participants, administer the intervention and monitor the participants for study outcomes and adverse events.

We are testing a real implementation of a new methodology for clinical trials, that we have called point-of-care clinical trials (POC-CTs), with features designed to maximize the pragmatic nature of studies. Aspects of the approach we describe here have been proposed or implemented by others [5–8] and discussed in detail under the name of the 'clinically integrated randomized trial' by Vickers and Scardino [9]. The defining characteristic here is that to the maximum extent possible the clinical trial apparatus is embedded in routine clinical care. Optimally, this would include recruitment and randomization of study subjects at their POC by their usual healthcare provider. Once randomized to a treatment arm subjects would continue to be treated by their healthcare provider with minimal or no deviation from usual care. Follow-up of participants would thus reflect current clinical practice. Assessment of subject compliance and practitioner adherence to protocol, and ascertainment of clinically relevant endpoints would be performed through medical record review, with minimal contamination of the clinical care 'ecosystem' by intrusive study dependencies. The intrusiveness of study operations, from randomization through endpoint ascertainment, would be greatly reduced if performed using tools familiar to healthcare providers and data already present in an electronic medical record (EMR).

A POC-CT shifts away from the asynchronous, distinct, and separate environments of research and clinical care, toward a real-time integrated system of research-based care. The goal of POC-CTs is to deliver the best care to patients while

learning from each experience and redefining that care. Under this new paradigm, ongoing results would be more rapidly and more likely adopted by providers who participated in the studies. By synthesizing research with practice and tools to learn from that process, participating facilities can move to the goal of becoming 'learning healthcare systems.'

In this article, we describe a specific POC-CT designed to test the feasibility and usefulness of the method, in answering a question of relevance to the Veterans Affairs (VA) Healthcare System. The clinical context and issues are described and ethical issues discussed. The use of outcome adaptive randomization to enhance implementation also addresses the frequentist operating characteristics of the design. The kinds of comparativeness questions best suited to POC-CT are argued.

Illustrative example: sliding scale insulin regimen *versus* weight-based insulin protocol

We describe a POC-CT which compares two common regimens of administering insulin therapy to hospitalized patients requiring insulin; the sliding scale and weight-based approach. The VA has an EMR that includes electronic ordering of medications and protocols for both of these insulin regimens. Review of EMR data at the VA Boston Healthcare System demonstrated that each of these two approaches is used with approximately equal frequency and discussions with treating clinicians indicated that choice of method administration is based on personal preference and not on patient specific determinants.

There are no published data comparing the effectiveness or the adverse effects of the sliding scale or a weight-based insulin protocol in treating inpatients with hyperglycemia. For the sliding scale, short acting insulin is administered three to four times daily according to the degree of hyperglycemia, and no basal insulin is administered. This regimen, therefore, responds to hyperglycemia after it occurs, and does not prevent it. The weight-based insulin protocol is a twice daily regimen of basal intermediate-acting insulin (NPH) plus a pre-meal twice a day regimen of short acting regular insulin, plus a correction dose of regular insulin depending on the degree of hyperglycemia. In addition, depending on the amount of the correction dose, the basal doses are adjusted upward for the next day's NPH insulin dose to manage the hyperglycemia.

Study design

Overall, the study is an open-label, randomized trial comparing sliding scale to a weight-based regimen in non-intensive care units (ICU) inpatients in a single large VA healthcare facility. There will be no modification to the treatment protocols already in use which will be accessed through the existing order entry menu. Consented patients will be randomized to treatment arms using an adaptive randomization method. Subjects are otherwise treated as usual. That is to say, there is no treatment protocol imposed other than insulin regimen beyond randomization. *There are no required diagnostic procedures and no study-specific follow-up events required*. Outcomes and covariates data will be collected directly from the computerized patient record system (CPRS). The primary endpoint is hospital length of stay (LOS); secondary endpoints include glycemic control and readmissions for glycemic control within 30 days of hospital discharge. Analysis will be based on intention to treat.

We considered using a cluster-randomized design, but the number of natural clusters (treatment units) within a hospital is small and having enough clusters to achieve adequate power would require opening the study at many hospitals, posing too many complex issues for a first use of POC-CT. Furthermore, we are interested in testing the feasibility of individual patient-level randomization, and the use of adaptive randomization to 'close the implementation gap.' While it is possible to imagine an adaptive cluster-randomized design, we have little information on the parameters necessary for design of such a study.

Eligibility

All non-ICU patients who require sliding scale or weight-based insulin therapy are eligible. The decision to obtain consent from a given individual will be made by the ordering clinician at the time of an insulin order (see section 'Methods'). There are no exclusions.

Treatment regimens

The treatment regimens are sliding scale and weight-based insulin as currently operationalized at the VA Boston Healthcare System. The ordering clinician finds these protocols under the electronic endocrine order menu and is led through order entry screens that insure standardization of the treatment protocol. The sliding scale and weight-based insulin regimens order menus in place at the

medical center were not modified other than to add a third choice allowing for randomization through the POC-CT mechanism.

Follow-up

Consenting subjects will be followed until 30 days of post-randomization. Following informed consent subjects will not be contacted by the study team either during their hospitalization or after discharge. All follow-up data will be collected *via* the EMR.

Data collection

Variables collected include demographics (age and gender); admission date, discharge date, and bed location (acute *vs*. non-acute); bed service (medical, surgical, and other); admission and other medical diagnoses (ICD-9 classification); glucose, blood counts, creatinine, and estimated glomerular filtration rate (GFR) values; and body temperature, medications, administered blood transfusion products, readmission date, and readmission diagnosis (ICD-9) if within 30 days of discharge. Non-VA hospitalization data for all subjects enrolled in Medicare will be available through a data-sharing agreement between VA and the Centers for Medicare & Medicaid Services.

Outcomes

The clinical outcomes of potential relevance that were considered included episodes of suspected hypoglycemia and measures previously used in studies examining potential benefit of improved glycemic control such as: (1) shortened length of hospital stay; (2) fewer infections; (3) fewer episodes of acute kidney injury; (4) less need for renal dialysis; (5) lower blood transfusion requirements; and (6) less neuropathy.

LOS is selected as the primary outcome, because LOS has important cost implications, lowers the risk of hospital-acquired complications including falls and infections, and might be expected to be shortened if diabetic control can be made more efficient. It is also readily ascertainable from the EMR. Secondary outcome measures include degree of glycemic control and readmission within 30 days of discharge with the primary readmission diagnosis of control of glycemia. Tertiary outcomes include infections, acute kidney injury, and anemia, all of which have been previously used as outcome measures in studies of insulin regimens. Infection will be defined as new antibiotic administration associated with either fever or leukocytosis. Acute kidney injury is defined as a decrease in estimated GFR of greater than 50% and anemia as a drop in the hemoglobin level of at least 2 g/dL.

Recruitment and enrollment

The POC-CT process is implemented using software tools available in CPRS. CPRS is the clinical care component of the Veterans Health Information Systems and Technology Architecture (VISTA), which supports clinical as well as administrative applications. Software tools available in CPRS include order sets (predefined customizable sets of orders), templates for clinical notes, decision logic (reminder dialog templates), and defined data objects that extract data from the medical record for display purposes (patient data objects). CPRS also has the ability to store flags (indicators in the data base) known as 'health factors' related to clinical parameters and flags derived from the ordering process. These tools make it possible to identify certain data elements in real time (e.g., an insulin order) and to incorporate programmatic logic into the medical record's workflow based on the value of data elements. The order sets and templates utilized for this project were designed to be consistent in format and process with the existing system.

The following describes the workflow of the study and demonstrates how CPRS processes already familiar to clinicians were adopted for POC-CT (Figures 1 and 2):

1) The VISTA order entry screen for insulin has been modified to include a third option in addition to the current options to order sliding scale or the weight-based regimen. The third option is labeled 'No preference for insulin regimen, consider enrollment in an inpatient study of Weight Based vs. Sliding Scale protocols' (Figure 3).
2) Clinicians who choose this third option will be presented with a brief description of the study and given the option to either proceed or not with consideration of their patient for study enrollment.
3) Clinicians who choose not to continue will click on the button labeled 'No. The patient may not be approached. Proceed with usual care.' and will be returned to the previous order entry screen to continue without further consideration of this trial.

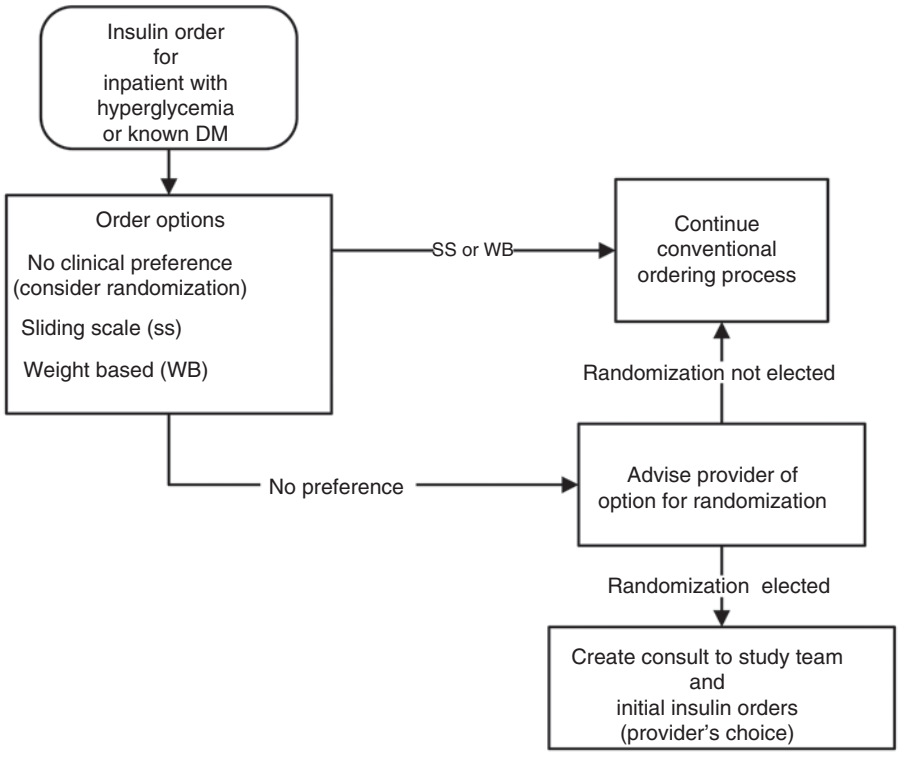


Figure 1 Initial order process performed by clinician

4) Clinicians who choose to proceed will click on the button labeled 'Yes. The research team may approach this patient for consideration of enrollment.' and will be brought to a consult entry screen. The consult entry screen will be pre-populated requesting a 'Research insulin dosing consent request.' After submitting this consult, the clinician will then be directed to the order entry menu and will order either sliding scale or weight-based insulin as per their choice. This order will serve as a holding order to provide insulin treatment until the patient can be consented and randomized.
5) Upon receiving the 'Research insulin dosing consent request,' the study nurse will discuss the study with the patient and obtain informed consent. If the patient declines enrollment, a template progress note completing the consult will be automatically entered. Patients who refuse randomization will be asked for consent to allow access to their VISTA data for comparison to the subset of patients who accepted randomization.
6) Patients who provide consent will be randomized through the VISTA system to one of the two insulin regimens. A template progress note activated by the study nurse will document randomization. This template progress note will generate 'health factors' that will serve to identify patients as subjects in the trial for tracking purposes in VISTA. It will also generate the order for whichever insulin regimen the subject was randomized to receive.
7) Progress notes (for both patients accepting and declining participation) and orders (for those accepting randomization) will be automatically forwarded to the original ordering clinician.
8) By signing these documents, the clinician completes the study enrollment process.

The protocol was approved by the VA Boston Institutional Review Board (IRB) who waived Health Insurance Portability and Accountability Act (HIPAA) authorization to allow the study team, once contacted and prior to seeing the patient, to have access to protected health information in the medical record. Importantly, clinicians, in simply referring patients to the study coordinator for recruitment and signing the insulin orders generated by the randomization procedures were not considered by the IRB to be 'engaged in clinical research' and thus were not required to be research credentialed.

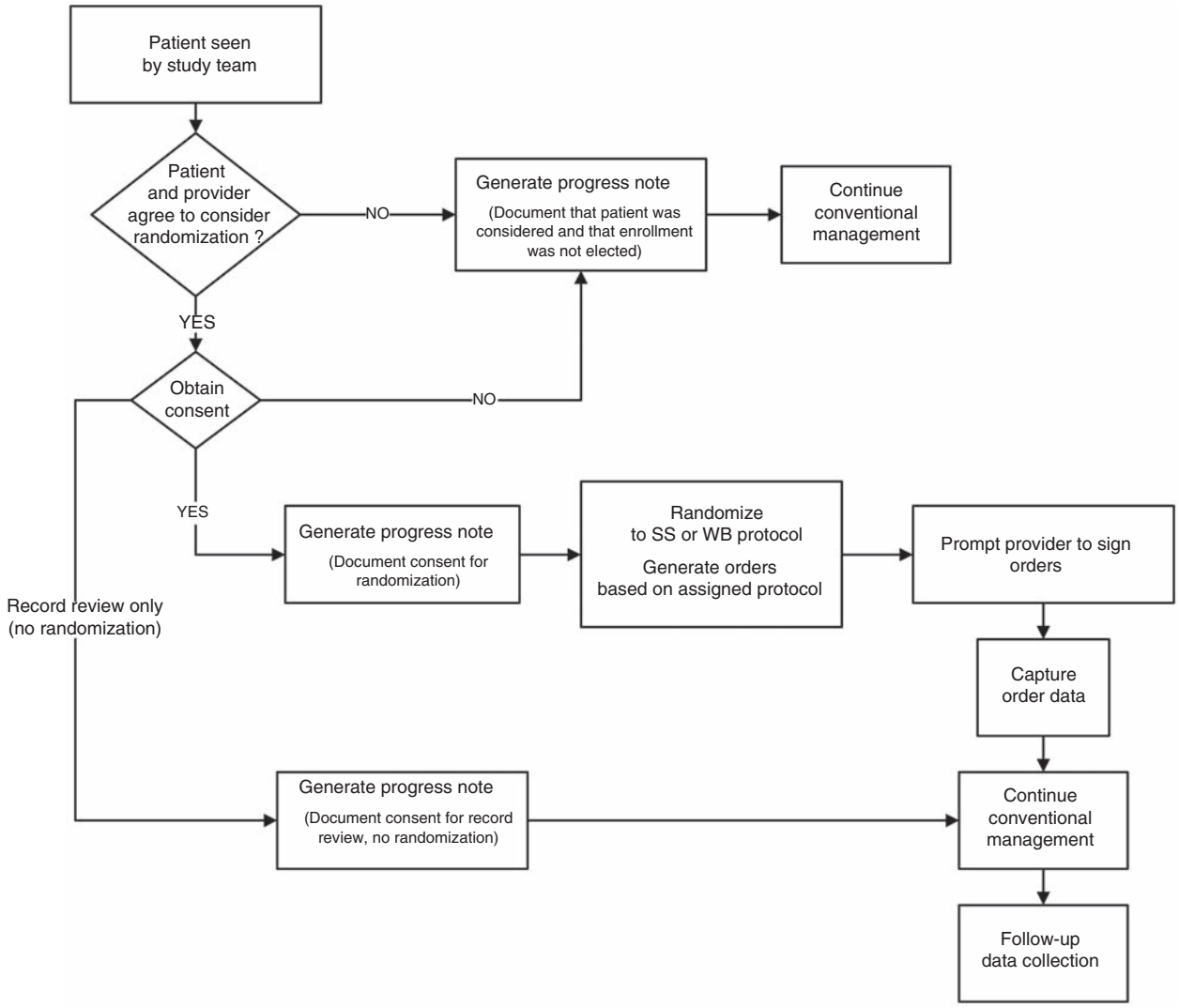


Figure 2 Workflow beginning when clinician has agreed to consider randomizing patient into one of two interventions

Statistical issues

We define three main aims: (1) to determine the physician and patient acceptance of POC randomization, (2) to test the null hypothesis of no difference against reasonable alternatives (two-sided), and (3) to demonstrate successful implementation of the superior strategy. The first aim requires descriptive statistical approaches, including estimating proportions and defining patient- and physician-level predictors of acceptance. The second aim requires tuning the design parameters to achieve acceptable operating characteristics. The third aim motivates an adaptive randomization, adjusting the assignment probabilities to increase the chances that patients are assigned to the better treatment.

Adaptive design

In the proposed study, the response or outcome is hospital LOS and the parameters of interest are the median LOS with each of the two protocols: (1) weight-based (Protocol A) and (2) sliding scale (Protocol B). We predict that the patients using the weight-based protocol will have a smaller median LOS than patients using the sliding scale protocol. To test this hypothesis, we propose using a Bayesian adaptive design.

The rules of adaptation considered herein modify the assignment probability each time the study accrues a new fixed number or 'batch' of patients, with practical batch sizes of at least 100 patients to allow more time for review and cleaning of data as is implicit in group sequential designs.

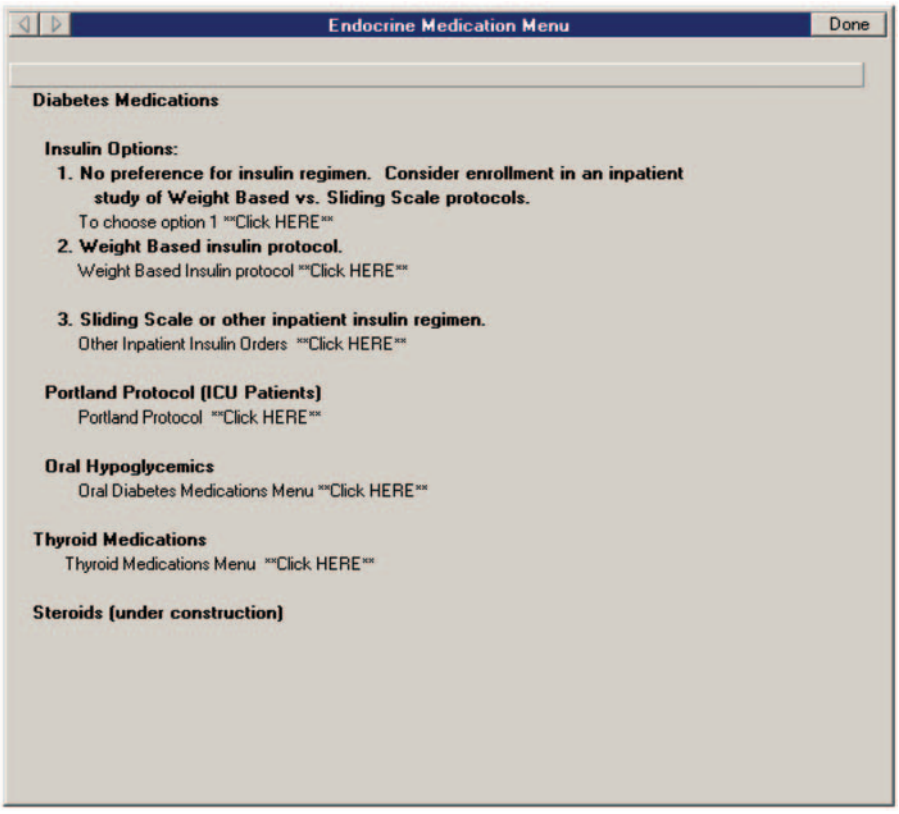


Figure 3 Screen shot of CPRS showing introduction of POC-CT option into the insulin options menu

According to this scheme (Figure 4)

1) First, subjects will be assigned to either weight-based protocol (Group A) with probability $\pi = 0.5$ or to sliding scale protocol (Group B) with probability $1 - \pi = 0.5$. This assignment probability is utilized for the first batch of patients.
2) Then, the data collected on the first group of subjects are used to calculate the probability that Protocol A is superior to Protocol B given the accumulated data, that is

$$\begin{aligned} p_A &= P(\text{Protocol A is superior to Protocol B}) \\ &= P(\theta_A < \theta_B | DATA) \end{aligned}$$

The '*DATA*' here refers to the data collected on the first batch of patients, with allowance for a period (UPDATE strip in Figure 4) in which the investigators clean the data and do the update and θ_A and θ_B are the median LOS in Groups A and B, respectively. The 'posterior' probability p_A ('*probability of Protocol A being superior to Protocol B given the data*') is calculated using Bayesian methods. Bayesian methods use prior information or beliefs, along with the current data, to guide the search for parameter estimates. Prior information/beliefs are input as a distribution, and the data then help refine that distribution and construct the posterior distribution. Our statistical model is based on an exponential data model for the LOS with conjugate Inverse Gamma prior for the median LOS [10]. Prior distributions in each group were chosen to be centered on the null median value and have a shape parameter α.

1) The posterior probability p_A is then used to evaluate whether the accumulated information overwhelmingly supports one protocol over the other so that the termination of the trial is warranted. In particular, we would stop the trial if

$$p_A > \kappa \text{ or } p_A < 1 - \kappa$$

where κ is the *cutpoint* reflecting the level of evidence demanded by the investigators to terminate the trial. If $p_A > \kappa$, then the study is terminated and Protocol A is chosen as being superior while if $p_A < 1 - \kappa$, the study is terminated and Protocol B is chosen to be superior. The value for κ is at the

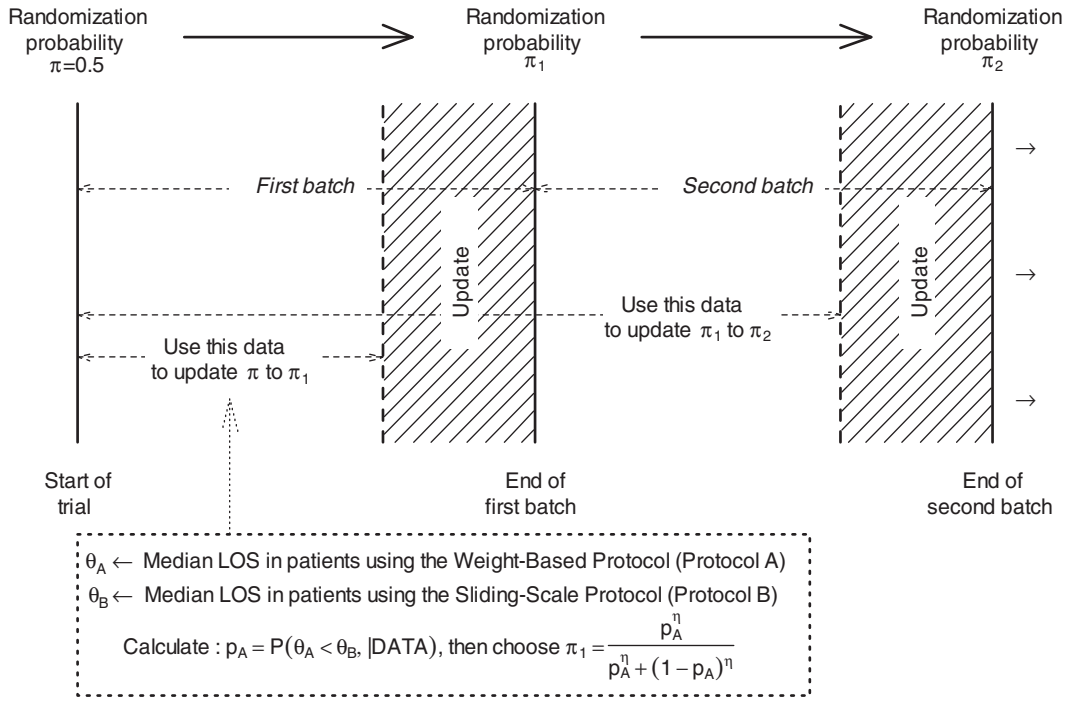


Figure 4 Diagram representing the flow of the design In the figure above, π represents the probability of assigning the weight-based protocol to a patient

investigators' disposal and it is usually a value that is close to 1 (for example 0.9, 0.95, or 0.99).

1) If the decision to terminate is not made, the posterior probability p_A is used to update the assignment probability to π_1 using the transformation [11]

$$\pi_1 = \frac{(p_A)^\eta}{(p_A)^\eta + (1-p_A)^\eta}$$

where $\eta > 0$ is a *calibration parameter*. If η is set to 1, the updated assignment probability is $\pi_1 = p_A$, while a value of $\eta = 0$ leads to a balanced randomization design. Values greater than 1 (less than 1) lead to more aggressive (less aggressive) adaptation.

1) The second batch of patients will then be assigned to Protocol A with probability π_1 and to Protocol B with probability $1 - \pi_1$. After the data on the second batch of patients are collected, the assignment probability π_1 is updated to π_2 using the above algorithm and the termination criterion is checked. If the termination criterion is met, the study is terminated. If not, the assignment probability π_1 is updated to π_2 using the above algorithm and the third batch is then enrolled.
2) This process is continued until either the termination criterion is met or the number of subjects enrolled reaches a pre-specified maximum number of subjects $N_{\max}$.

Proposed design

Extensive computer simulations were done to select a design for the study based on their operating characteristics. The following operating characteristics were considered in selecting the final design:

1) *Overall Type I error* – the chance of declaring one of the two protocols better at any time during the trial when in fact there is no difference between the two protocols.
2) *Overall power* – the chance of declaring a protocol better at any time during the trial when in fact that protocol is better.
3) *The number of patients assigned to each protocol.* The number of patients enrolled will depend on the data collected and hence is a random variable.
4) *Time until a decision is made.* The duration of the study will depend on the data collected and hence is a random variable.

We chose a design with the following parameters: prior shape parameter $\alpha = 100$, batch size = 200, cutpoint $\kappa = 0.99$, calibration parameter $\eta = 0.5$, and maximum number of patients to be randomized $N_{max} = 3000$. In addition, the updation occurs after 150 patients of each batch have entered the study, we do not update or allow stopping after the first batch, and we censor the LOS at 30 days.

We studied the above design under various scenarios. Our null hypothesis is that the median LOS with both protocols is 5 days. As alternative, we posit a minimal clinically important reduction of at least 12% in median LOS.

The operating characteristics of the design are represented in Table 1.

Type I error: Under the assumption of no difference (first row in Table 1 – median LOS is 5 days with both protocols) the probability of (incorrectly) selecting either protocol as superior was 0.06.

Power: Under the alternatives (median LOS with Protocol A < median LOS under Protocol B) presented in the remaining rows of the table, the probability of correctly selecting Protocol A represents the power. For a difference of 12% in median LOS, across the interim looks, the design will correctly select Protocol A as superior with 86% probability (power), while the probability of wrongly selecting Protocol B as superior decreases fast to levels close to 0%. The decision to stop increases with time (Figure 5); thus, the probability or terminating the trial by the 6th interim look (after 1400 subjects have been enrolled) is 50% and it increases to 86% by the 14th look (after all 3000 subjects have been enrolled).

From among the many alternatives designs we evaluated, we briefly discuss here the *balanced* design that has the same parameters as the design presented above. Additional information on the simulation study including the R [12] script used in running the simulations can be obtained from the authors.

With a balanced design, the Type I error is the same, the power is slightly higher (for example, 77% *vs*. 71% to detect a difference with Protocol A of 10% in median LOS), the median number of patients enrolled is about the same (~2000),

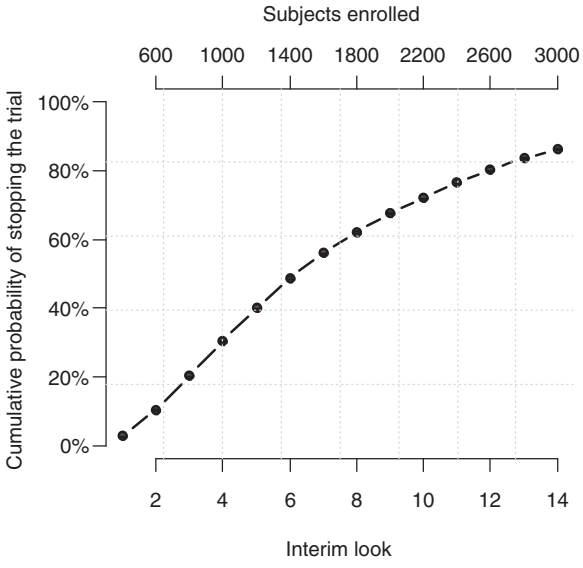


Figure 5 Cumulative probability of stopping the trial across interim looks; assumed median LOS with Protocols B and A are 5 and 4.4 days, respectively

Table 1 Operating characteristics of the proposed design

Difference in median LOS (B–A) in days [median under Protocol B = 5 days]	Probability of selecting Protocol A as superior (%)	Probability of selecting Protocol B as superior (%)	Median number of patients on Protocol A	Median number of patients on Protocol B	Median duration (days)[a]
0	3	3	1495	1461	599
0.1	8	1	1634	1292	598
0.2	17	0	1738	1125	597
0.3	30	0	1791	969	595
0.4	51	0	1719	778	581
0.5	71	0	1434	598	408
0.6	86	0	1075	465	316
0.7	95	0	825	380	240
0.8	99	0	673	332	201
0.9	100	0	540	289	164
1	100	0	506	268	157

[a]In calculating the duration of the study, we assumed an accrual rate of 5 patients per day.

Table 2 Operating characteristics under lognormal data model

Difference in median LOS (B–A) in days [median under Protocol B = 5 days]	Probability of selecting Protocol A as superior (%)	Probability of selecting Protocol B as superior (%)	Median number of patients on Protocol A	Median number of patients on Protocol B	Median duration (days)[a]
0	4	3	1469	1473	599
0.1	8	2	1594	1317	599
0.2	16	1	1711	1163	597
0.3	28	0	1759	998	595
0.4	46	0	1724	832	587
0.5	62	0	1600	696	485
0.6	78	0	1244	535	360
0.7	90	0	924	414	275
0.8	96	0	715	352	210
0.9	99	0	626	309	193
1	100	0	522	278	160

[a]In calculating the duration of the study, we assumed an accrual rate of 5 patients per day.

however, while with the balanced design the enrollment is balanced, with our proposed design the number of patients assigned to the superior treatment is higher.

The operating characteristic simulation is dependent on the accuracy of the data model used to generate the LOS. In Table 1, we use the exponential model to generate the data, as well as to do the updating. Thus, it makes the assumption that the Bayesian model is correctly specified, as is done in most published work, when estimating (frequentist) operating characteristics. But the LOS data from a historical sample of patients approximating the proposed study intake criteria indicates a heavier tail, such as log-normal. Therefore, we assessed the sensitivity of the assumptions by using the log-normal model to generate the data (but still using the exponential model for the updates; Table 2).

The difference between these two simulations illustrates the modest sensitivity of the operating characteristics to misspecification of the data model. For example, the Type I error estimate rises from 6% to 7%, and the power at a difference of 0.5 days drops from 71% to 62%. However, we consider the Type I error less relevant in this context, comparing the effectiveness of two widely used procedures for setting dose. In a different context, the Type I error might be more important. The probability of making the right choice when it matters (a full day difference) is high (100%) in the log-normal scenario, too. These results illustrate the value of a hybrid approach, where the Bayes method is confined to updating the randomization probability (thus closing the implementation gap and maximizing the number of patients receiving the right treatment) and inference is based on operating characteristics from a range of more realistic models.

Discussion

POC-CT methodology is well suited for studies with the following features:

- Interventions already approved by the FDA.
- A clinical question where there is equipoise regarding clinically relevant alternative interventions.
- Interventions that are part of routine practice, well tolerated, and have well-recognized toxicities which mitigates the need for adverse event monitoring beyond that in routine clinical care.
- Subject identification, inclusion and exclusion criteria, and endpoints that are accurately obtained from the EMR.
- Outcomes are objective and require little or no adjudication.
- Study protocol requiring minimal deviations from usual care.
- No systematic laboratory or clinical follow-up required for either safety or comparative effectiveness.

This trial is designed to be on the pragmatic extreme of the clinical trial spectrum with the subject consent process being the sole perturbation of the clinical care 'ecosystem.' The absence of study specific interventions, procedures, and monitoring together with passive data capture attempts to maximize the relevance of the findings to

current practice at the VA Boston Healthcare System. Adaptive randomization is designed to assign subjects preferentially to the treatment arm that, in real time, appears superior, with an 'efficacy' stopping rule that has acceptable Type I error. If the study terminates without reaching its 'efficacy' boundary, it will reliably rule out a substantial difference, in which case cost, convenience, and other factors will dictate which treatment arms continue to be supported. Such direct translation of study results into clinical practice defines a 'learning healthcare system.'

The clinical question posed in this protocol, comparison of insulin administration methods, was chosen because it is amenable to a maximally pragmatic study as defined by the PRECIS criteria and because:

- Broad participation by healthcare providers is expected. The clinical question is compelling and in practice there is apparent equipoise between the two regimens in that roughly half of patients are currently treated by each technique.
- The inclusion/exclusion criteria will allow enrollment of nearly all the VA Boston patients who require the intervention.
- The study interventions are currently utilized at VA Boston, have known toxicities that are monitored as part of usual care, and thus require no specific study related monitoring.
- All study data elements are objective, resident in the EMR and do not require study specific interactions or visits for capture.
- Adaptive randomization methodology leads to real-time incorporation of study results into practice, if one treatment proves superior.

The ability to implement this study is made possible by the VA's EMR environment. CPRS is in use at all the VA's 1500-plus points of care and was designed to incorporate clinical data as part of efforts to improve clinical care. As a result, it features several packages that allow end users to automatically generate reports, 'listen' for certain values associated with patient data objects, consider these values with programmatic logic, and introduce information and workflows directly into the EMR. To capitalize on this level of flexibility, most VA healthcare systems employ Clinical Application Coordinators, who use these tools to create and report measures of the quality of care, to implement guidelines, and to create clinical reminders based on the priorities of each hospital. This infrastructure will allow for the relatively easy roll-out of this and other POC-CT studies system-wide as well as systematic implementation of findings.

The ability to use existing functionalities, as opposed to developing custom software is important for a number of reasons. First, development of new software functionality is constrained by time for development, testing, and approval, and development resources. Second, by capitalizing on existing system functionality, we increase the likelihood of a successful deployment to other VA hospitals or clinics, each one of which employs CPRS. Finally, although this particular use of CPRS may be novel, the POC-CT processes are presented through familiar interfaces and into a culture of robust CPRS use, which we hope will facilitate adoption of this approach.

The ability of institutions to implement POC-CTs is dependent on the ability to use the EMR to: (1) identify events as they present in real time; (2) intervene in the clinical care workflow; and (3) track longitudinal data. It is worth noting that these functionalities are critical to the creation and implementation of many novel approaches to learn from and improve healthcare based on real data and that few systems offer such capabilities to end users. The need for such functionalities is of particular relevance in light of the US Federal Government's upcoming investment of $19 billion to support the adoption of EMRs [13]. Much of this funding is contingent on the adoption of 'certified' EMR systems and the 'meaningful use' of such systems. Definitions that require flexible integration with EMR data and workflows are essential to meeting the goals of such enormous investments [14].

The ethical and practical considerations of informed consent have been extensively discussed and debated [15–19] as have methods such as cluster randomization which might obviate or preclude individual informed consent [20,21]. Detailed analyses of these considerations are outside the scope of this article. However, as POC-CTs or similarly designed trials become an important component of clinical research, it will be incumbent on investigators, ethicists, and IRBs to fully consider the potential benefits and apparently minimal incremental risks of a POC-CT, and to take responsibility for helping their healthcare systems to lower the barriers to successful study design and implementation of improvements in care.

A study coordinator will obtain written informed consent for all subjects entered into this trial. This requirement accounts for a significant proportion of the study cost and introduces the single most tangible perturbation to the usual care workflow. We recognize that replacement of such full written informed consent by an alternative (such as simple 'notification' by the healthcare provider and verbal consent by the subject with subsequent

randomization through a fully automated computerized process) would result in an even more efficient design, with a closer match to clinical care. The IRB could consider such a variation on the usual research informed consent, on a study-by-study basis, especially when the POC-CT results in care materially identical to usual clinical practice. Parallel requirements would be a waiver of HIPAA authorization to obtain study data from the EMR and acknowledgement that treating clinicians who authorize automated randomization are not 'engaged' in research.

A POC-CT will likely require significantly less study-specific infrastructure and cost than traditional RCTs (after the up-front investment in coordinating center and informatics, already made by the VA). These advantages together with an economy of scale once an investment in the methodology has been made could lead to low incremental cost per study as well as allowing study designs of sufficient duration to capture clinically relevant (as opposed to surrogate) endpoints.

Limitations

Several issues may impede adoption of POC-CTs. Some patients may find it surprising and troubling that healthcare providers do not know what is the best treatment for them. This disclosure could make the consent process lengthy and difficult. Although the medical community might be at equipoise regarding treatment options, individual healthcare providers may have strong treatment preferences, either in general or for particular individual patients. Both of these issues could have ramifications for recruitment rates and the success of a POC-CT. We note that 'reluctance to randomize' is an issue for all RCT designs, not just POC-CT.

Most (if not all) uses of POC-CT we envision would have an open (unblinded) design, which raises the possibility of cross-contamination of treatments, or differential clinical interventions due to physicians' perceptions of patients' needs, or other failures of the exclusion principle, such as observational bias in the outcome. Therefore, the use of POC-CT may be restricted to clinical situations where the effects are likely to be minimal. We think that the EMR-based protocols we compare here, as well as the outcome of LOS, sharply reduce physician unblinding as a threat. We emphasize that POC-CT is not a universal alternative to the classical double-blind RCT with its many controls for bias; rather, it can be seen as a competitor to observational studies, by removing the particular bias from selection by indication that plagues such non-experimental studies.

Our pragmatic intent requires us to rely on individual clinician judgment of eligibility, which is another mark of distinction between POC-CT and conventional trials, which often have elaborate procedures for defining 'inclusion and exclusion.' This certainly restricts the use of POC-CT to contexts where such precision is unnecessary. However, it also contributes to the 'ecological validity' of treatment effects.

Highly pragmatic POC-CTs such as this study may yield results that are locally convincing but are not easily generalized to other healthcare systems. A healthcare system such as the VA, motivated to conduct POC-CTs and with the organization and infrastructure capable of supporting it, could generate 'locally selfish' evidence-based medicine to gain evidence of comparative effectiveness most relevant to its population and systems. In general, comparative effectiveness findings are most applicable to the systems and individuals who participated in its creation rather than to the 'free riders' – those who may desire evidence-based medicine but who are unwilling to be a part of that evidence.

The above may suggest that the POC-CT approach is limited to a narrow range of clinical questions and contexts. We are just now beginning to expand our list of possible use cases, and we do not want to speculate in advance of the facts. We agree with Vickers and Scardino [9] that features of POC-CT might be implemented in practice in four distinct areas: surgery, 'me too' drugs, rare diseases, and lifestyle interventions. In addition to questions of optimizing care (such as the insulin example described here) use cases currently under consideration include technology introduction (imaging, robotics, and biomarker-guided therapy), pre-hydration with bicarbonate *versus* saline with or without n-acetylcysteine in contrast-induced nephropathy, and comparing prolonged exposure and cognitive processing therapies as alternative treatment strategies for post-traumatic stress disorder.

Finally, the proposed study design using outcome adaptive randomization leads to real-time implementation into practice, and stimulates reconsideration of the role of the traditional peer review process that subjects study results to expert outside review before planning their implementation in practice.

Funding

This study was supported by the Department of Veterans Affairs Cooperative Studies Program. Dr Lavori's efforts have been supported by grants UL1

RR025744 and R01 MH051481 to Stanford University.

References

1. **Schwartz D, Lellouch J.** Explanatory and pragmatic attitudes in therapeutic trials. *J Chron Dis* 1967; **20**: 637–48.
2. **Schwartz D, Flamant R, Lellouch J.** *Clinical Trials*. London, Academic Press, 1980.
3. **Tunis SR, Stryer DB, Clancy CM.** Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *J Am Med Assoc* 2003; **290**: 1624–32.
4. **Thorpe KE, Zwarenstein M, Oxman AD, *et al*.** A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009; **62**: 464–75.
5. **Chalmers TC.** Randomization of the first patient. *Med Clin N Amer* 1975; **59**: 1035–38.
6. **Baum M.** New approach for recruitment into randomized clinical trials. *Lancet* 1993; **1**: 812–13.
7. **Luce BR, Kramer JM, Goodman SN, *et al*.** Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Ann Intern Med* 2009; **151**: 206–09.
8. **Zwarenstein MF, Dainty KN, Quan S, *et al*.** A cluster randomized trial evaluating electronic prescribing in an ambulatory care setting. *Trials* 2007; **8**: 28.
9. **Vickers AJ, Scardino PT.** The clinical-integrated randomized trial: proposed novel method for conducting large trials at low cost. *Trials* 2009; **10**: 14.
10. **Giles FJ, Kantarjian HM, Cortes JE, *et al*.** Adaptive randomized study of idarubicin and cytarabine versus troxacitabine and cytarabine versus troxacitabine and idarubicin in untreated patients 50 years or older with adverse karyotype acute myeloid leukemia. *J Clin Oncol* 2003; **21**: 1722–27.
11. **Thall PF, Wathen JK.** Practical Bayesian adaptive randomization in clinical trials. *Eur J Cancer* 2007; **43**: 859–66.
12. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing [document on the Internet]. The Institute for Statistics and Mathematics, Vienna, 2009. Available at: http://www.R-project.org (accessed 9 June 2010).
13. American Medical Association. American Recovery and Reinvestment Act of 2009 (ARRA), Summary of Major Health Care Provisions, 2009.
14. **D'Avolio L.** Electronic medical records at a crossroads; impetus for change or missed opportunity. *J Am Med Assoc* 2009; **302**: 1109–10.
15. **Baum M.** Do we need informed consent? *Lancet* 1986; **2**: 911–12.
16. **Edwards SJL, Lilford RJ, Braunholtz, *et al*.** Ethical issues in the design and conduct of randomized controlled trials. *Health Technol Assess* 1998; **2**(15): i–vi, 1–132.
17. **Emanuel EJ, Wendler D, Grady C.** What makes clinical research ethical? *J Am Med Assoc* 2000; **283**: 2701–11.
18. **Marco CA.** Impact of detailed informed consent of research subjects' participation: a prospective, randomized trial. *J Emerg Med* 2008; **34**: 269–75.
19. **Parvizi J, Chakravarty R, Og B, Rodriguez-Paez A.** Informed consent: is it always necessary? *Injury* 2008; **39**: 651–55.
20. **Hutton JL.** Are distinctive ethical principles required for cluster randomized controlled trials? *Stat Med* 2001; **20**: 473–88.
21. **Sabin JE, Mazor K, Meterko V, Goff SL, Platt R.** Comparing drug effectiveness at health plans: the ethics of cluster randomized trials. *Hasting Cent Rep* 2008; **38**: 39–48.

By Marshall Fleurant, Rachel Kell, Jennifer Love, Chelsea Jenter, Lynn A. Volk, Fang Zhang, David W. Bates, and Steven R. Simon

DOI: 10.1377/hlthaff.2010.1020
HEALTH AFFAIRS 30,
NO. 7 (2011): 1256–1264

Massachusetts E-Health Project Increased Physicians' Ability To Use Registries, And Signals Progress Toward Better Care

Marshall Fleurant (mfleuran@post.harvard.edu) is a general internal medicine research fellow in the Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, in Boston, Massachusetts.

Rachel Kell is the evaluation coordinator at the Massachusetts eHealth Collaborative, in Waltham.

Jennifer Love was formerly a research assistant at Partners HealthCare System, in Boston.

Chelsea Jenter is a research project manager at the Brigham and Women's Hospital, in Boston.

Lynn A. Volk is associate director of the Partners HealthCare System.

Fang Zhang is an instructor in population medicine at the Harvard Pilgrim Health Care Institute.

David W. Bates is division chief, general internal medicine, at the Brigham and Women's Hospital, and a professor of health policy and management at the Harvard School of Public Health, in Boston.

Steven R. Simon is chief of the Section of General Internal Medicine, Veterans Affairs Boston Healthcare System.

ABSTRACT The ability to generate and use registries—lists of patients with specific conditions, medications, or test results—is considered a measure of physicians' engagement with electronic health record systems and a proxy for high-quality health care. We conducted a pre-post survey of registry capability among physicians participating in the Massachusetts eHealth Collaborative, a four-year, $50 million health information technology program. Physicians who participated in the program increased their ability to generate some types of registries—specifically, for laboratory results and medication use. Our analysis also suggested that physicians who used their electronic health records more intensively were more likely to use registries, particularly in caring for patients with diabetes, compared to physicians reporting less avid use of electronic health records. This statewide project may be a viable model for regional efforts to expand health information technology and improve the quality of care.

The American Recovery and Reinvestment Act of 2009, which included the Health Information Technology for Economic and Clinical Health (HITECH) provisions, allocated more than $48 billion to promote the spread of health information technology (IT) throughout the United States.[1,2] Widespread use of electronic health records is seen as a foundation for health reform.[3] However, considerable doubt remains as to how deeply health IT will penetrate US health care and whether it will actually produce the anticipated quality improvements.

To foster the adoption of health IT among ambulatory practices, HITECH authorizes the Office of the National Coordinator for Health Information Technology to establish regional extension centers to assist providers in selecting and implementing certified interoperable electronic health records for their practices. These centers aim to disseminate "lessons learned" and "best practices" throughout their communities and to promote participation in health information exchange.[4]

Policy makers and leaders of regional extension centers will naturally look to existing models of communitywide efforts to promote the adoption of health IT. One of the most visible examples of such programs is the Massachusetts eHealth Collaborative, a statewide consortium of health care stakeholders founded in 2004 to improve the quality and safety of health care through community-based adoption of health IT.[5–10]

During 2006–08, the Massachusetts eHealth Collaborative sponsored a program to implement electronic health records within ambulatory medical practices and establish health information exchange. With a $50 million grant from Blue Cross and Blue Shield of Massachusetts, the collaborative used a competitive

process to select three communities to participate in the program. Additional details about the program are presented elsewhere.[5,7]

Our study evaluates one aspect of the potential impact of this program on the quality of health care. Because effects on patient outcomes and costs of care are likely to require several years to become measurable, we identified the ability to generate patient registries as an early proxy measure for the Massachusetts eHealth Collaborative's potential impact on the quality of care.

Registry capability, defined as the ability to generate lists of patients based on defined clinical characteristics requiring specific action, is frequently viewed as an essential tool for improving the health care of individuals and populations.[11,12] It is included in the "meaningful use" criteria of the Centers for Medicare and Medicaid Services.[13] Although registries can be paper based, practices with electronic health records are considerably more likely than practices that use paper records to have registry capability.[11]

Therefore, we hypothesized that the Massachusetts eHealth Collaborative would increase registry capability among participating practices. As a secondary hypothesis, we examined whether primary care practices with more extensive use of their electronic health records were more likely to use their registries in the care of patients with chronic conditions, as compared with practices with less avid use of electronic health records.

Study Data And Methods

STUDY DESIGN We evaluated the implementation of electronic health records using pre- and post-intervention surveys to measure physicians' perceived ability to generate registries. The Partners HealthCare Human Research Committee approved the study protocol.

INTERVENTION The Massachusetts eHealth Collaborative installed robust electronic health records and provided work-flow redesign and technical support at no cost to the offices of participating physicians during 2006–08. The electronic health records were certified by the Certification Commission for Health Information Technology. "Practice consultants" with expertise in implementing ambulatory electronic health records and redesigning office practices met with office staff and physicians in preparation for the record systems' deployment, during the deployment, and after the systems were fully deployed.

The intervention was evaluated according to several measures, including the use of technology, assessment of barriers and facilitators, implementation tactics, safety, impact on quality of care, and fiscal parameters.[6,7,10]

SETTING AND PARTICIPANTS The Massachusetts communities of Brockton, Newburyport, and North Adams were selected for the pilot program. The characteristics of these communities and the physicians and practices within them reflected the demographics and practice characteristics of physicians and practices across Massachusetts.[8,14]

A total of 167 physician practices participated, representing 86 percent of all eligible primary care and specialty practices in these communities. All were invited to respond to pre-intervention (2005) and post-intervention (2009) surveys (see the online Appendix).[15]

For the pre-intervention survey we identified 464 physicians from 167 practices; 355 completed the survey (response rate: 77 percent).[14] In 2009, 468 physicians were eligible to participate; of these, 319 completed the survey (response rate: 68 percent).

Between 2005 and 2009, some practices dissolved, and some physicians departed, while new physicians entered the communities. A total of 163 physicians from 134 practices completed both the 2005 and 2009 survey questionnaires. This set of physicians constituted the main analytic sample.

SURVEY DESIGN The pre- and post-intervention surveys were based on similar statewide surveys of physicians, described elsewhere.[11,14,16–18] Briefly, the 2005 survey measured physicians' attitudes toward the use of computers in health care and specifically assessed the ability of each practice to generate registries. The post-intervention (2009) survey retained the original survey items and added new questions related to health information exchange and the electronic health record implementation process.

MAIN OUTCOME MEASURES The main outcome measure was the ability to generate registries, as reflected by the responses to the following questions asked identically in 2005 and 2009: "With your current medical record system (paper and/or electronic), how easy would it be for you or your staff to generate the following information about your patients? A) List of patients by diagnosis or health risk (e.g., diabetes); B) List of patients by laboratory results (e.g., patients with abnormal hematocrit levels); C) List of patients by medications they currently take (e.g., patients on warfarin)." These questions referred to the three different types of registries we referenced in the study: for diagnoses, laboratory results, and medications.

Responses were elicited along a five-point Likert-type scale and were dichotomized to classify physicians as able (very easy, somewhat easy, somewhat difficult, very difficult) versus

not able (cannot generate) to generate each registry type. Because "very difficult" and "cannot generate" may reflect a similar set of physicians who need health IT implementation support, we dichotomized the outcome in post hoc analyses to classify physicians as capable (very easy, somewhat easy, somewhat difficult) versus needing support (very difficult, cannot generate).

OTHER OUTCOME MEASURES The 2009 survey asked physicians about their actual use of registry functions in office practice. Specifically, it asked whether their office, by either paper or electronic means, did the following tasks annually for patients with diabetes and patients with coronary artery disease: prompted the practice to notify patients who are overdue for office visits; generated a list of patients who are overdue for tests; and generated a list of patients with clinical data suggesting that they needed an intervention (for example, their blood tests indicated elevated hemoglobin A1c greater than 7, or their systolic blood pressure measured greater than 140).

OTHER VARIABLES The surveys also recorded physician characteristics, such as age, sex, race, years in practice, and number of outpatients seen per week, and practice characteristics, such as number of physicians in the practice, specialty, ownership, and financial resources available for expansion.

DATA ANALYSIS To assess for nonresponse bias, we compared demographics and practice characteristics of physicians who completed both the 2005 and 2009 surveys with those of physicians who completed only the 2005 (baseline) survey.

For our initial pre-post analysis, we compared the proportion of physicians who were able to generate each type of registry in 2009 versus 2005. We first examined these proportions among all survey respondents in 2009 ($N = 319$) compared with all survey respondents in 2005 ($N = 355$) to provide a "snapshot" of registry capability in each time period. In all subsequent analyses, we restricted the sample to physicians who had completed both the 2005 and 2009 surveys ($n = 163$).

We compared pre- and post-intervention rates separately for each of the main outcome measures of ability to create registries—based on diagnoses; laboratory results; and medications—using McNemar's test for all three outcomes. We also used logistic regression with generalized estimating equations to generate confidence intervals and odds ratios and to account for repeated measures (that is, the same physician's completing the survey in 2005 and 2009). In all analyses, results were qualitatively similar between generalized estimating equations and McNemar's test.

The ability to generate registries is an integral component of the patient-centered medical home.

In a secondary cross-sectional analysis, we examined the relationship between the use of electronic health record functions and registry functions among internal medicine and family medicine physicians ($n = 87$) in 2009, for whom the measure was most relevant. We classified physicians as "high" and "low" electronic health record users based on a previously developed measure of usage.[18] This aggregate measure reflects each physician's reported use of ten key functions in the electronic health record, with each physician's calculated score ranging from 0 (no usage of available functions) to 1 (consistent usage of all available functions).

We dichotomized the sample as high users (above the median value of 0.8) or low users (below the median). This score did not include any measure of registry usage. We compared rates of use of each of the three registry types (diagnosis, laboratory test, medication) among high and low electronic health record users, using Fisher's exact test. Analyses were performed using the statistical analysis software package SAS, version 9.2.

STUDY LIMITATIONS Our study had several limitations. Most important, our study lacked a control group, which prevented us from excluding longer-term trends as an explanation for our observed findings. It is possible that a trend toward increased attention to quality measurement and improvement during the study period, including interest in pay-for-performance initiatives, might have included registry capabilities. Our unpublished statewide data, mentioned in the Discussion section, do in fact suggest a noncyclical trend toward increased registry capability. However, the magnitude of the pre-post intervention effect observed in this study seems too great to be accounted for by this secular trend alone.

Another important limitation is that we re-

stricted our main analyses to physicians who completed both pre- and post-intervention surveys. Although this approach minimizes potential confounding, it may limit generalizability. Future studies and similar intervention programs must consider the inevitable flux of physicians into and out of communities.

We also note that the study evaluated the effect of the Massachusetts eHealth Collaborative on registry capabilities through physician self-report rather than actual use. All physicians within the collaborative actually had the ability to generate registries as part of their electronic health records, whether or not they were aware of it, because the installed electronic health records all featured registry capability. Future studies should measure actual registry usage from data extracted directly from electronic health records, as will ultimately be required to determine "meaningful use."[19] Furthermore, we did not assess whether traditional quality measures or patient health outcomes improved during the study period.

Study Results

BASELINE CHARACTERISTICS Baseline characteristics of the 163 physicians completing both the 2005 and 2009 surveys are shown in Exhibit 1. These physicians were similar to those who responded to only the 2005 survey ($n = 192$) with respect to age ($p = 0.32$), sex ($p = 0.74$), years in practice ($p = 0.10$), specialty ($p = 0.34$), average number of patients seen per week ($p = 0.66$), and resources available for practice expansion ($p = 0.06$). However, they were more likely to practice in groups of more than five physicians (58 percent versus 46 percent; $p = 0.01$) and more likely to be full or partial owners of their practices (66 percent versus 58 percent; $p = 0.02$).

MAIN OUTCOME MEASURES Compared with all survey respondents in 2005, all survey respondents in 2009 were more likely to be able to generate each registry type (Exhibit 2).

Among the 163 physicians who completed both surveys, the ability to generate a diagnosis registry was similar over the study period (88 percent in 2009; 89 percent in 2005). In contrast, 78 percent of these physicians in 2009 were able to generate a laboratory results registry, compared with 44 percent in 2005, and 83 percent in 2009 could generate a medication registry, compared with 33 percent in 2005.

In analyses that adjusted for all available covariates, physicians in 2009 were more likely to be able to generate a laboratory and a medication registry than they were in 2005. In contrast,

EXHIBIT 1

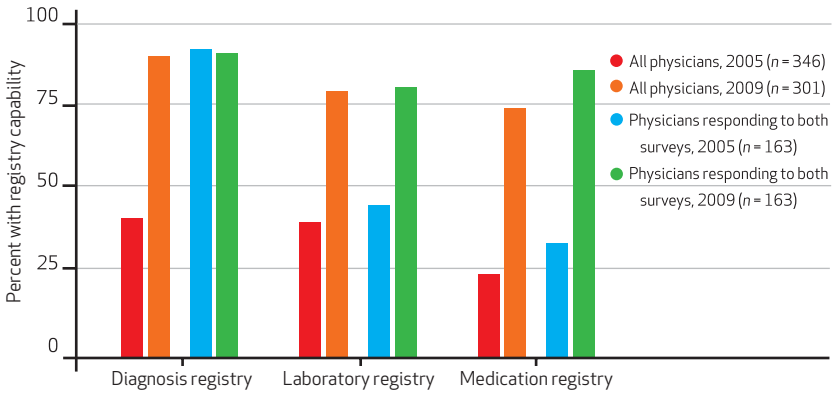
Baseline Characteristics Of The Massachusetts eHealth Collaborative Physicians And Practices

Characteristic	Participants (*N* = 163)
Age, mean, in years (SD)	49 (9.9)
Male	123 (75%)
Race	
White	130 (80%)
Asian American	15 (9%)
Hispanic	4 (2%)
African American	4 (2%)
Other/mixed	7 (4%)
Years with current practice group, median (IQR)	9 (3–18)
Practice owner[a]	108 (66%)
Specialty	
Primary care[b]	78 (48%)
Non–primary care[c]	85 (52%)
Practice size	
1–2 physicians	41 (25%)
3–5 physicians	28 (17%)
>5 physicians	94 (58%)
Number of outpatients per week, median (IQR)	90 (50–119)
Adequate resources for practice expansion, number[d]	30 (18%)

SOURCE Authors' data. **NOTES** All measures are based on 2005 survey data, with the exception of race, which was ascertained only in 2009. SD is standard deviation. IQR is interquartile range. [a]Practice owner indicates physicians who described themselves as full or part owners of the practice. [b]Primary care includes general internal medicine ($n = 26$), pediatrics ($n = 16$), family practice ($n = 25$), and primary care (not otherwise specified; $n = 11$). [c]Non–primary care includes general surgery and surgical subspecialties ($n = 39$), subspecialties of internal medicine ($n = 21$), obstetrics/gynecology ($n = 12$), and other specialties ($n = 13$). [d]Participants indicated whether their practice had adequate resources for improvement or expansion.

EXHIBIT 2

Ability Of Physicians In The Massachusetts eHealth Collaborative To Generate Diagnosis, Laboratory Test, And Medication Registries



SOURCE Authors' data. **NOTE** The exhibit shows, from left to right for each registry type, the proportion with registry capability among all physicians in 2005, all physicians in 2009, physicians in 2005 who completed both surveys, and those in 2009 who completed both surveys.

there was no change in ability to generate a diagnosis registry (odds ratio: 1.0; 95 percent confidence interval: 0.4–2.4) (Exhibit 3). The online Appendix shows complete results of the multivariate analysis.[15]

Practice size was significantly associated with the ability to generate all registry types: Smaller practices were consistently less likely than larger practices to be able to generate registries. Specialty type was significantly associated with the ability to generate both medication and laboratory registries, but not diagnosis registries. Older physicians were less likely to be able to generate diagnosis registries compared to younger physicians; however, a significant age effect was not observed with medication or laboratory registries (Exhibit 3).

In a post hoc analysis, with the registry capability outcome dichotomized as those who are capable of generating registries and those who need support, we found that the Massachusetts eHealth Collaborative intervention increased capability for all three registry types: diagnosis (odds ratio: 1.9; 95% confidence interval: 1.0–3.9); laboratory results (odds ratio: 7.6; 95% confidence interval: 4.2–13.7); and medication (odds ratio: 10.2; 95% confidence interval: 5.2–9.9).

ELECTRONIC HEALTH RECORD AND REGISTRY USE We were interested in primary care physicians' actual use of registries. Eighty-seven internal medicine and family medicine physicians completed the 2009 questionnaire. We dichotomized these physicians as "high users" ($n = 44$) and "low users" ($n = 43$) of the electronic health record.

Exhibit 4 shows the proportion of these physicians who indicated that they performed registry-based tasks for patients with diabetes mellitus and coronary artery disease. For diabetes, there was a trend toward greater use of registry functions among high users. The result reached statistical significance for the use of registry functions to remind patients about overdue testing (odds ratio: 2.8; 95% confidence interval: 1.1–7.1). There was no consistent relationship observed between electronic health record usage and the use of registry functions for patients with coronary artery disease.

Discussion

During 2006–08 the Massachusetts eHealth Collaborative implemented electronic health records and provided work-flow redesign and technical support to more than 160 physician prac-

EXHIBIT 3

Correlates Of Physician Practices' Ability To Generate Registries, Massachusetts eHealth Collaborative

	Odds ratio		
Variable	**Diagnosis registry**	**Laboratory registry**	**Medication registry**
Intervention[a]	1.0	7.2**	13.9****
Primary care[b]	1.1	0.4**	2.2**
Practice size			
1–2 physicians	0.1**	0.6	0.4
3–5 physicians	0.2**	0.4**	0.2**
>5 physicians (referent)	—	—	—
Age >50	0.8	1.5	1.7

SOURCE Authors' data. **NOTES** Adjusted for age, sex, race, years in practice, ownership, practice volume, and financial resources. Odds ratios and confidence intervals for all variables included in the multivariate model are in the online Appendix; see Note 15 in text. [a]Intervention indicates the pre-post effect of the Massachusetts eHealth Collaborative intervention program. [b]Primary care includes general internal medicine, pediatrics, and family practice. **$p < 0.05$ ****$p < 0.001$

EXHIBIT 4

Performance Of Registry Tasks Among High And Low Users Of The Electronic Health Record (EHR) In The Massachusetts eHealth Collaborative

	For diabetes mellitus			For coronary artery disease		
Generate list of patients with	**High EHR users (*n* = 43)**	**Low EHR users (*n* = 43)**	***p* value**	**High EHR users (*n* = 41)**	**Low EHR users (*n* = 43)**	***p* value**
Overdue visits	60%	51%	0.39	51%	47%	0.67
Overdue tests	55	30	0.02	34	30	0.70
Abnormal labs	51	37	0.20	28	35	0.47

SOURCE Authors' data. **NOTES** *N* = 87. Of the eighty-seven primary care physicians, one did not provide responses to the items related to diabetes mellitus, while three did not provide responses to the items regarding coronary artery disease. We classified physicians as high and low EHR users. We calculated an EHR usage score for each physician that ranged from 0 (no usage of available functions) to 1 (consistent usage of all available functions), and we dichotomized the sample as high users (above the median value of 0.8) or low users (below the median). This usage score did not include any measure of registry usage.

tices in a well-funded community-based outreach intervention. In this pre-post analysis, we found that the intervention was associated with improved ability to generate registries based on laboratory test results and medications, but not based on diagnosis. The failure to observe an increase in diagnosis registry capability is probably due to the fact that nearly 90 percent of practices reported being able to generate diagnosis registries prior to the intervention. Practices may have been able to generate diagnosis registries from computerized practice management systems—used for billing purposes—which may have been implemented well before this study. These practice management systems would not have enabled the practices to generate laboratory results or medication registries.

The results of this study suggest that communitywide implementation of electronic health records through an organized approach of strategic planning, work-flow redesign, and technology deployment may increase capacity to use registries—a proxy measure of the ability to deliver high-quality health care.

POTENTIAL FOR WIDER IT IMPLEMENTATION Much of the current multibillion-dollar federal investment in health information technology is expected to take the form of community-based interventions to increase electronic health record adoption, such as through the Beacon Communities and regional extension centers. The Massachusetts eHealth Collaborative's efforts provide a prototype, as evident in the increased registry capability from 2005 to 2009. They support the notion that these programs can achieve widespread health IT implementation and potentially improve the quality of care delivered.

Further study is needed to determine whether practices in other states have similar levels of registry capability. The health IT regional extension centers should assess the various registry capabilities of their constituent practices to tailor the transformation efforts needed for each practice.

OTHER REASONS FOR IMPROVEMENT Although these results show that implementation of this program led to improvements in registry capability, alternative explanations need to be considered. This study was not designed to measure whether or not the quality of care actually improved as a result of the Massachusetts eHealth Collaborative program. Most notably, it is conceivable that the pre-post increases in registry capability could be due to secular trends, and not a result of the program intervention.

Over a similar time period in a companion study, randomly sampled physicians across Massachusetts reported increased capability for generating registries, but the magnitude of the increases (absolute increases of 4–10 percent) were considerably smaller than those seen in the present study (33 percent increase for laboratory result registry capability and 50 percent increase for medication registry capability) (Adam Wright, Brigham and Women's Hospital; personal communication, August 3, 2010).

PRACTICE SIZE The observation that 89 percent of practices at baseline were able to generate registries deserves further mention. More than 70 percent of large physician organizations can generate diagnosis registries.[20] However, little is known about small and medium-size practices, such as those in our study. Anecdotally, physicians in the study communities validated our study's estimates of nearly universal registry capability; however, physicians also noted wide variability in the extent to which their practices actually use registries.

SPECIALTY Based on our results, the relationship between specialty and registry capability is unclear. Specialists were more likely to be able to generate laboratory registries, while primary care physicians were more likely to be able to generate medication registries. There was no re-

lationship between specialty and diagnosis registry capability. Some specialists, such as nephrologists and oncologists, whose practices rely heavily on the interpretation of laboratory test results, might logically be expected to have developed robust capability to generate laboratory registries to support their practice needs. Whether this type of practice specialization accounts for the differences observed between specialists and primary care doctors remains unknown. Moreover, it is difficult to explain why primary care physicians should be more likely to be able to generate medication registries than specialists, many of whom prescribe medications with the same frequency as generalists. Clearly, further study is needed to characterize how different specialists and primary care physicians use registries.

Specialists increasingly appreciate the value of regional and national registries that report complications and other measures of the quality of their care.[21,22] Moreover, physicians in all specialties are recognizing the need to track and manage their own patient populations over time.[23] However, little is known about how specialists and primary care physicians use registries, so further research is warranted.

CORRELATION WITH ELECTRONIC HEALTH RECORD USE Our secondary analyses suggested a trend toward greater reported registry use among physicians with greater electronic health record usage, although this finding was seen for diabetes care but not for coronary artery disease. Diabetes mellitus is more prevalent than coronary artery disease and may be more widely recognized as a target for population health management and quality improvement than is the case for coronary artery disease.

It is possible that in the Massachusetts effort we studied, practice consultants may have emphasized the potential value of using registries for diabetes care as an example of how the electronic health record could be used to improve quality of care. Moreover, because diabetes is more prevalent than coronary artery disease, it is plausible that practices aggressively using their electronic health records would embrace registry usage for diabetes more than for coronary artery disease.

These results support the model of planned, organized, community-based health IT implementation.

Although not conclusive, the results of this analysis are consistent with the notion that greater electronic health record usage may be associated with greater registry usage. This study also adds to the literature in its evaluation of a broad community implementation using commercial electronic health record systems.

Studies of registry use and capability to date have been generally descriptive and cross-sectional.[24,25] This analysis provides evidence that the intervention resulted in improved registry capability—an important indicator of health care quality. The ability to generate registries is also an integral component of the patient-centered medical home.[26]

POLICY IMPLICATIONS Although our results must be interpreted with some caution, this study nonetheless has important implications for health care policy. These results support the model of planned, organized, community-based health IT implementation programs. They suggest that Beacon Community programs and regional extension centers may be successful, although both approaches should be evaluated. We found that a community-level intervention was associated with increased ability to generate registries, which should ideally result in commensurate improvement in the quality, safety, and cost of health care. ■

An earlier version of this article was presented in oral abstract form at the Society of General Internal Medicine Regional meeting in Providence, Rhode Island, March 2010, and as a poster at the Society of General Internal Medicine national meeting in Minneapolis, Minnesota, April 2010. This study was funded by the Agency for Healthcare Research and Quality, Cooperative Agreement No. 1UC1HS015397, and by the Massachusetts eHealth Collaborative. Marshall Fleurant is supported by a Ruth L. Kirschstein T32 Training Grant (No. T32 HP12706). The authors thank Adam Wright for his helpful input.

NOTES

1 Steinbrook R. Health care and the American Recovery and Reinvestment Act. N Engl J Med. 2009; 360(11):1057–60.

2 Recovery Accountability and Transparency Board. Funding for Recovery Act innovation and technology [Internet]. Washington (DC): The Board; 2010 [cited 2010 Aug 30]. Available from: http://www.recovery.gov/News/featured/Pages/InnovationandTechnology

Funding.aspx
3 Taylor R, Bower A, Girosi F, Bigelow J, Fonkych K, Hillestad R. Promoting health information technology: is there a case for more-aggressive government action? Health Aff (Millwood). 2005;24(5):1234–45.
4 Office of the National Coordinator for Health Information Technology. HITECH programs: Health Information Technology Extension Program [Internet]. Washington (DC): Department of Health and Human Services; 2011 [last updated 2011 Jan 27; cited 2011 May 24]. Available from: http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__hitech_programs/
5 Goroll AH, Simon SR, Tripathi M, Ascenzo C, Bates DW. Community-wide implementation of health information technology: the Massachusetts eHealth Collaborative experience. J Am Med Inform Assoc. 2009;16(1):132–9.
6 Rudin RS, Simon SR, Volk LA, Tripathi M, Bates DW. Understanding the decisions and values of stakeholders in health information exchanges: experiences from Massachusetts. Am J Public Health. 2009;99(5):950–5.
7 Mostashari F, Tripathi M, Kendall M. A tale of two large community electronic health record extension projects. Health Aff (Millwood). 2009;28(2):345–56.
8 Tripathi M, Delano D, Lund B, Rudolph L. Engaging patients for health information exchange. Health Aff (Millwood). 2009; 28(2):435–43.
9 DeBor G, Diamond C, Grodecki D, Halamka J, Overhage JM, Shirky C. A tale of three cities—where RHIOs meet the NHIN. J Health Inf Manag. 2006;20(3):63–70.
10 Halamka J, Aranow M, Ascenzo C, Bates D, Debor G, Glaser J, et al. Health care IT collaboration in Massachusetts: the experience of creating regional connectivity. J Am Med Inform Assoc. 2005;12(6): 596–601.
11 Wright A, McGlinchey EA, Poon EG, Jenter CA, Bates DW, Simon SR. Ability to generate patient registries among practices with and without electronic health records. J Med Internet Res. 2009;11(3):e31.
12 Asch SM, Baker DW, Keesey JW, Broder M, Schonlau M, Rosen M, et al. Does the collaborative model improve care for chronic heart failure? Med Care. 2005;43(7):667–75.
13 Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010;363(6):501–4.
14 Simon SR, Kaushal R, Jenter CA, Volk LA, Burdick E, Poon EG, et al. Readiness for electronic health records: comparison of characteristics of practices in a collaborative with the remainder of Massachusetts. Inform Prim Care. 2008;16(2): 129–37.
15 To access the Appendix, click on the Appendix link in the box to the right of the article online.
16 Simon SR, Kaushal R, Cleary PD, Jenter CA, Volk LA, Poon EG, et al. Correlates of electronic health record adoption in office practices: a statewide survey. J Am Med Inform Assoc. 2007;14(1):110–7.
17 Quinn MA, Wilcox A, Orav EJ, Bates DW, Simon SR. The relationship between perceived practice quality and quality improvement activities and physician practice dissatisfaction, professional isolation, and work-life stress. Med Care. 2009; 47(8):924–8.
18 Simon SR, Kaushal R, Cleary PD, Jenter CA, Volk LA, Orav EJ, et al. Physicians and electronic health records: a statewide survey. Arch Intern Med. 2007;167(5):507–12.
19 Medicare and Medicaid programs; electronic health record incentive program: final rule. Fed Regist. 2010 Jul 28;75(144):44313–588.
20 Rittenhouse DR, Shortell SM, Gillies RR, Casalino LP, Robinson JC, McCurdy RK, et al. Improving chronic illness care: findings from a national study of care management processes in large physician practices. Med Care Res Rev. 2010(67): 301–20.
21 Haggstrom DA, Clauser SB, Taplin SH. The health disparities cancer collaborative: a case study of practice registry measurement in a quality improvement collaborative. Implement Sci. 2010;5(42):1–15.
22 Schmittdiel J, Bodenheimer T, Solomon NA, Gillies RR, Shortell SM. Brief report: the prevalence and use of chronic disease registries in physician organizations: a national survey. J Gen Intern Med. 2005; 20(9):855–8.
23 Casalino LP, Rittenhouse DR, Gillies RR, Shortell SM. Specialist physician practices as patient-centered medical homes. N Engl J Med. 2010;362(17): 1555–8.
24 Wang CJ, Marken RS, Meili RC, Straus JB, Landman AB, Bell DS. Functional characteristics of commercial ambulatory electronic prescribing systems: a field study. J Am Med Inform Assoc. 2005;12(3): 346–56.
25 Kaelber D, Greco P, Cebul RD. Evaluation of a commercial electronic medical record (EMR) by primary care physicians 5 years after implementation. AMIA Annu Symp Proc. 2005:1002.
26 Bates DW, Bitton A. The future of health information technology in the patient-centered medical home. Health Aff (Millwood). 2010;29(4): 614–21.

ABOUT THE AUTHORS: MARSHALL FLEURANT, RACHEL KELL, JENNIFER LOVE, CHELSEA JENTER, LYNN A. VOLK, FANG ZHANG, DAVID W. BATES & STEVEN R. SIMON

Marshall Fleurant is a general internal medicine research fellow at the Harvard Pilgrim Health Care Institute and Harvard Medical School.

Marshall Fleurant and coauthors report in this month's *Health Affairs* on results from their survey of physicians who participated in the Massachusetts e-Health Collaborative. The study reports that participating physicians increased their ability to generate registries of laboratory results and medications—essential tools for improving health care of individuals and populations.

Fleurant is a general internal medicine research fellow in the Department of Population Medicine at the Harvard Pilgrim Health Care Institute and Harvard Medical School. His current interests are in health information technology and its effect on

primary care and ambulatory practice. He received a medical degree from the Albert Einstein College of Medicine and a master of public health degree from the Harvard School of Public Health.

Rachel Kell is the evaluation coordinator at the Massachusetts eHealth Collaborative.

Rachel Kell works for the Massachusetts eHealth Collaborative, where she has been involved in the evaluation of its electronic health record implementation pilot program. She earned a master of public health degree in health policy and management from Boston University.

Jennifer Love is a former research assistant at Partners HealthCare System.

Jennifer Love is a former research assistant with the Clinical and Quality Analysis IT group at Partners HealthCare System. She is now a medical student at the Perelman School of Medicine, University of Pennsylvania.

Chelsea Jenter is a research project manager at Brigham and Women's Hospital.

Chelsea Jenter is a senior research project manager in the Division of General Medicine at Brigham and Women's Hospital. She earned a master of public health degree in health policy and management from Boston University.

Lynn A. Volk is an associate director at Partners HealthCare System.

Lynn A. Volk is associate director of the Clinical and Quality Analysis Department within Partners HealthCare System's Information Services. She earned a master's degree in health science from the Johns Hopkins Bloomberg School of Public Health.

Fang Zhang is an instructor in population medicine at the Harvard Pilgrim Health Care Institute.

Fang Zhang is director of programming in the Drug Policy Research Group at the Harvard Pilgrim Health Care Institute and an instructor in population medicine at Harvard Medical School and Harvard Pilgrim. He has a master's degree in statistics from Worcester Polytechnic Institute and a doctorate in biostatistics from Boston University.

David W. Bates is division chief, general internal medicine, at Brigham and Women's Hospital.

David Bates is chief of the Division of General Internal Medicine at Brigham and Women's Hospital, a professor of medicine at Harvard Medical School, and a professor of health policy and management at the Harvard School of Public Health, where he codirects the Program in Clinical Effectiveness. His medical degree is from Johns Hopkins.

Steven R. Simon is chief of the Section of General Internal Medicine, Veterans Affairs Boston Healthcare System.

Steven Simon is chief of the Section of General Internal Medicine at the Veterans Affairs Boston Healthcare System and an associate professor at Harvard Medical School and Brigham and Women's Hospital. He is a general internist and health services researcher with a focus on interventions to improve the quality and safety of health care, principally through the use of health information technology and education on the evaluation of innovative medical education and quality improvement programs.

Simon received a medical degree from Yale University and a master of public health degree from the Harvard School of Public Health.

Patient Navigation: The Promise to Reduce Health Disparities

Karen M. Freund, MD, MPH

Women's Health Unit, Section of General Internal Medicine, Evans Department of Medicine, and Women's Health Interdisciplinary Research Center, Boston University School of Medicine, Boston, MA, USA.

J Gen Intern Med 26(2):110–2
DOI: 10.1007/s11606-010-1593-5

There is a great need for effective interventions to address health disparities for vulnerable populations, which may be defined by low education, low literacy, low income, inadequate health insurance, or minority racial / ethnic status. Many efforts have focused on the social determinants of health, resting on the notion that the magnitude of health outcome disparities affecting so many different racial and ethnic groups argues against biologic differences, but is rather due to social and cultural influences on how patients use or receive care. This implies that personalized medicine to address genetic risk alone will not eliminate health disparities, but rather systems interventions to improve access and the process of care delivery are critical to improving the health of the entire nation.

Looking specifically at cancer outcomes, lower screening rates are well described for underserved populations[1]. Even where screening rates are similar between racial and ethnic groups,[2] differences exist in follow up rates after abnormal screening,[3] resulting in differences in stage and size of tumors at diagnosis which in turn contribute to ongoing disparities[4]. Safety net institutions, which are less likely to have sufficient resources to track and support patients, have populations with greater barriers to completing their diagnostic or treatment care. This perfect storm of patients with limited resources to support themselves and their families when cancer is diagnosed, and resource-poor safety net institutions with limited resources to provide extra support is at the heart of many of our nation's health disparities.

This problem has become more acute with the economic downturn, where resources to safety net institutions are increasingly strained, even while the numbers of patients losing employer-based coverage grows. As an example, in Massachusetts, the two largest safety net hospitals, Boston Medical Center, and Cambridge Health Alliance have experienced huge budget gaps. This is due to both marked state reductions in their Medicaid reimbursement rates, and payment rates for insured patients from health insurance companies being sometimes two to three times lower than those provided to nearby teaching hospitals [5–7]. As a result, our reimbursement systems perpetuate the continued presence of fewer resources to institutions whose patient populations have few resources.

Published online December 15, 2010

Added to this disparity the advent of pay for performance,[8] where institutions will receive payments based upon performance to quality benchmarks for their entire populations. While most would agree that paying for improved outcomes provides the appropriate incentives to our health care systems, it also serves to put at greater risk those institutions caring for those most in need[9]. Those institutions that require additional resources to provide care may lose additional resources to care systems with patient populations better able to adhere to their doctors' recommendations. How to make the playing field level for safety net institutions is not obviously clear. The notion of having a lower achievement standard to receive performance payments for safety net institutions on its face seems to codify rather than improve health disparities, and it also is based on the notion that disparities are an intractable problem.

Patient navigation is an emerging model of improving healthcare delivered to vulnerable populations, which has primarily focused on improved cancer outcomes. Patient navigation has been defined as the logistic and emotional support offered to persons through the cancer care continuum from screening, through diagnostic evaluation and cancer treatment. The goal of patient navigation is to support patients in overcoming logistical barriers to care and facilitate timely access to quality cancer care that meets cultural needs for all patients. Navigators work to address financial and insurance issues, coordinate appointments and care among multiple cancer providers, address language and health literacy needs and train patients to advocate for themselves in the health care system[10]. The term "patient navigation" was coined by Dr. Harold Freeman, as a care management system to address cancer disparities. His initial demonstration project[11] has been followed by studies demonstrating that additional staffing resources to providers in safety net institutions can improve intermediate outcomes of completion and timeliness of screening and diagnostic care, as a mechanism to ultimately improving health outcomes[12–14].

Patient navigation shares many attributes with other care management models. Most patient navigation programs have been housed directly with the providers of health care and facilitate care through providing a liaison between patients and the health care team. A recent Centers for Medicare and Medicaid (CMS) program investigating the clinical and financial benefit of care management programs funded multiple models of care, ranging from off-site telephone-based interventions, to clinically-based programs linked with the providers of care. Of the 15 funded programs, only two showed a benefit; both were programs with face to face as opposed to telephone-only contact, and were programs housed and organized with the providers of care[15].

The two studies published in this issue of JGIM provide further support of the patient navigation model. Both programs are funded through the Avon Foundation, which has

provided consistent support for breast cancer care to safety net programs over the past decade. The study by Phillips and colleagues[16] addressed low HEDIS-measured screening mammography, and studied the impact of patient navigation through a pre-post difference in difference analysis. The 10% improvement in HEDIS rates achieved by the intervention is noteworthy in that it would allow such a safety net system to benefit from most pay for performance programs. The findings illustrate that unstable housing and incorrect contact information in their patient populations remain limitations to this patient navigation model.

Donelan and colleagues[17] studied patient satisfaction with the patient navigator model of care. They compared patients from their community health center affiliates, predominantly low income, minority women, with a primarily white, educated population whose primary care was not through a safety net institution. Both populations reported similar satisfaction with care after abnormal breast cancer screening. The lack of demographic overlap between the two groups prevents multivariable or propensity adjustment to fully understand what impact the navigation model played in the results. The authors conclude that the navigated and non-navigated groups report similar perceptions of the quality of their care. It is not clear whether this indicates that the navigated group would have been equally satisfied with their care without navigation or whether prior to navigation, the minority population would have reported a poorer experience with care that was ameliorated with navigation. Further research needs to study this issue with appropriate control populations, to understand if patient navigation is serving to benefit the quality of care in low income and minority communities.[18,19]

The initial reports suggest that patient navigation holds promise to address care needs to vulnerable populations, and bridge the disparities gap. Currently, however, patient navigation systems are more likely to be found and marketed in systems caring and targeting patients who are insured, employed and educated. Paradoxically, if patient navigation is an effective modality for improving care, there is a risk of it increasing rather than eliminating health disparities. Few safety net institutions have the ongoing resources to support this augmentation to care, and insurers do not reimburse this care. This is in contrast to hospitals and health care systems providing care to predominantly insured populations, where resources are present to support patient navigation services. The websites of most private cancer care centers tout the benefits of their patient navigation, and these are now seen as standards of care by some cancer care accreditation organizations including the National Accreditation Program for Breast Centers[20], and the National Comprehensive Cancer Networks[21]. Even as patient navigation is studied to ensure appropriate care to the underserved, its major implementation has been in insured populations who already have better health outcomes. Health care reform has at its goal a transition into a care management approach, as a necessary step to improving quality and reducing cost. We will need to carefully watch that resources including patient navigation and case management are provided to safety net institutions to address the challenges that their populations face, if we hope to reduce and not increase disparities in care.

Acknowledgements: *The author has received support from the Avon Foundation, Komen for the Cure Foundation, and is Principal Investigator of one of the NCI funded Patient Navigation Research Program grants U01 CA116892.*

Disclaimer: *Editor's note: Dr. Freund is a co-author on the paper by Phillips et al. (reference #16).*

Conflicts of Interest: *None Disclosed.*

Corresponding Author: *Karen M. Freund, MD, MPH; Women's Health Unit, Section of General Internal Medicine, Evans Department of Medicine, and Women's Health Interdisciplinary Research Center, Boston University School of Medicine, 801 Massachusetts Avenue, Suite 470, Boston, MA 02118, USA (e-mail: Karen.Freund@bmc.org).*

REFERENCES

1. **Weir HK, Thun MJ, Hankey BF, Ries LAG, Howe HL, Wingo PA, Jemal A, Ward E, Anderson RN, Edwards BK.** Annual report to the nation on the status of cancer, 1975-2000, featuring the uses of surveillance data for cancer prevention and control. J Natl Cancer Inst. 2003;95(17):1276–99.
2. **Green EH, Freund KM, Posner MA, David MM.** Pap smears rates among Haitian immigrant women in eastern Massachusetts. Public Health Rep. 2005;120:133–139.
3. **Battaglia TA, Santana MC, Bak S, et al.** Predictors of timely follow-up after abnormal cancer screening among women seeking care at urban community health centers. Cancer. 2010;116(4):91321.
4. **Richards WA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ.** Influence of delay on survival in patients with breast cancer: a systematic review. Lancet. 1999;353(9159):1119–26.
5. **Krasner J.** State urged to review fees to elite hospitals. The Boston Globe [Internet]. Available at: http://www.boston.com/news/health/articles/2008/11/20/state_urged_to_review_fees_to_elite_hospitals. Accessed 10 November 2010.
6. **Allen S, Bombardieri M.** An Unhealthy System; Part 1: A Healthcare System badly out of Balance. Available at: http://www.boston.com/news/local/articles/2008/11/16/a_healthcare_system_badly_out_of_balance. Accessed 12 November 2010.
7. **McNamara E.** The Patrick Administration likes the notion of universal health coverage. It just doesn't want to pay for it. Boston Magazine [Internet]. . Available at: http://www.bostonmagazine.com/articles/universal_health_coverage. Accessed 10 Novemeber 2010.
8. **Rosenthal MB, Frank RG, Li Z, Epstein AM.** Early experience with pay-for-performance. From concept to practice. JAMA. 2005;294:1788–1793.
9. **Hong CS, Atlas SJ, Change Y, et al.** Relationship between patient panel characteristics and primary care physician clinical performance rankings. JAMA. 2010;304:1107–1113.
10. **Freund KM, Battaglia TA, Calhoun E, et al.** National Cancer Institute Patient Navigation Research Program: methods, protocol, and measures. Cancer. 2008;113(12):3391–3399.
11. **Freeman HP.** Patient navigation: A community centered approach to reducing cancer mortality. J Cancer Educ. 2006 Spring;21(1 Suppl):S11-4.
12. **Ell K, Vourlekis B, Lee PJ, Xie B.** Patient navigation and case management following an abnormal mammogram: A randomized clinical trial. Prev Med. 2007;44(1):26–33.
13. **Battaglia TA, Roloff K, Posner MA, Freund KM.** Improving follow-up to abnormal breast cancer screening in an urban population. A patient navigation intervention. Cancer. 2007;109(2 Suppl):359–67.
14. **Lasser KE, Murillo J, Medlin E, et al.** A multilevel intervention to promote colorectal cancer screening among community health center patients: results of a pilot study. BMC Fam Pract. 2009;10:37.
15. **Peikes D, Chen A, Schore J, Brown R.** Effects of care coordination on hospitalization, quality of care, and health care expenditures among Medicare beneficiaries. JAMA. 2009;301(6):603–618.
16. **Phillips CE, Rothstein JD, Beaver K et al.** Patient navigation to increase mammography screening among inner city women. J Gen Intern Med. 2010. doi:10.1007/s11606-010-1527-2.
17. **Donelan K, Mailhot JR, Dutwin D et al.** patient Perspectives of clinical care and patient navigation in follow-up of abnormal mammography. J Gen Intern Med. 2010 Jul. doi:10.1007/s11606-010-1436-4.

18. **Ferrante JM, Chen PH, Kim S.** The effect of patient navigation on time to diagnosis, anxiety, and satisfaction in urban minority women with abnormal mammograms: a randomized controlled trial. J Urban Health. 2008;85(1):114–24.
19. **Jean-Pierre P, Fiscella K, Freund KM et al.** Structural and reliability analysis of a patient satisfaction with cancer-related care measure: a Multisite patient navigation research program study. Cancer. 2010 Oct 4.
20. National Accreditation Program for Breast Centers. Program standards/components. Available at: http://accreditedbreastcenters.org/standards/standards.html. Accessed 12 November 2010.
21. National Comprehensive Cancer Network. The case manager or patient navigator: providing support for cancer patients during treatment and beyond. Available at: http://www.nccn.com/index.php/life-with-cancer/understanding-treatment/152-case-managers-for-cancer-patients. Accessed 12 November 2010.

AIDS PATIENT CARE and STDs
Volume 25, Number 3, 2011

DOI: 10.1089/apc.2010.0104

Incidence and Predictors of Acute Kidney Injury in an Urban Cohort of Subjects with HIV and Hepatitis C Virus Coinfection

Shikha Garg, M.D., M.P.H.,[1] Melanie Hoenig, M.D.,[2] Erika M. Edwards, Ph.D., M.P.H.,[3] Caleb Bliss, M.A.,[4] Timothy Heeren, Ph.D.,[4] Sheila Tumilty, R.N.,[5] Alexander Y. Walley, M.D., M.Sc.,[6] Margaret J. Koziel, M.D.,[1] Paul R. Skolnik, M.D.,[7] C. Robert Horsburgh, M.D., M.U.S.,[3,5] and Deborah Cotton, M.D., M.P.H.[3,5]

Abstract

Coinfection with hepatitis C (HCV) significantly increases the risk of acute and chronic renal disease in HIV-infected individuals. However, the burden of acute kidney injury (AKI) directly attributable to HIV among HCV-infected individuals and associated risk factors are not well understood. Within a prospective cohort, AKI episodes were identified by a rise in creatinine of 0.5 mg/dL. Incidence of first AKI events was calculated for HIV/HCV coinfected versus HCV monoinfected subjects, and multivariable analyses using Cox proportional hazards were performed to identify predictors of AKI. Throughout the study period, 35% HIV/HCV coinfected and 17% HCV monoinfected subjects developed AKI, with incidence of 8.74/100 person-years and 3.53/100 person-years, respectively (hazard ratio (HR) 2.48; [95% confidence interval (CI) 1.50, 3.74]). In multivariable analysis, HIV coinfection (HR 2.19 [1.33, 3.62]), decompensated cirrhosis (HR 6.64 [3.81, 11.6]), and cocaine use (HR 2.06 [1.15, 3.71]) were independently associated with AKI. HCV genotype, HCV viral load, hazardous drinking, and heroin use were not associated with AKI. Study limitations included potential misclassification bias of HCV-infected individuals as serial HIV antibody testing was not routinely performed after study entry, and inability to adjust for tenofovir use in multivariable analysis. In conclusion, among subjects with HCV infection, decompensated cirrhosis, HIV coinfection, and cocaine use are associated with increased risk of AKI. These findings highlight the importance of preventing and treating cirrhosis, controlling HIV coinfection, and reducing cocaine use in HIV/HCV coinfected persons.

Introduction

Renal disease is an increasingly important cause of morbidity and mortality among HIV-infected individuals.[1,2] It is estimated that 30% of HIV-infected patients in the United States have abnormal renal function.[1] Recent epidemiologic studies have revealed that coinfection with hepatitis C virus (HCV) confers an even greater risk of both acute kidney injury (AKI) and chronic kidney (CKD) disease among individuals infected with HIV.[3–7]

AKI is a common complication among both ambulatory and hospitalized HIV-infected patients in the highly active antiretroviral therapy (HAART) era.[3,4] In one study involving hospitalized HIV patients by Wyatt and colleagues,[3] AKI was associated with a 5.83 increased odds of in-hospital mortality. Franceschini and colleagues[5] examined predictors of AKI among HIV-infected patients and found that coinfection with HCV was significantly associated with AKI, along with male gender, CD4 count less than 200 cells/mm^3, HIV viral load (VL) greater than 10,000 copies per milliliter, and HAART exposure. A recent meta-analysis of studies involving HIV/HCV coinfection and renal disease found a pooled relative risk of 1.64 [95% confidence interval (CI) 1.21, 2.23] for AKI in HIV/HCV-coinfected versus HIV-monoinfected patients. However, the authors commented that few of the studies provided a clear definition of HCV coinfection, and only one

[1]Section of Infectious Diseases, [2]Section of Nephrology, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts.
[3]Department of Epidemiology, [4]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts.
[5]Section of Infectious Diseases, [6]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts.
[7]Center for HIV/AIDS Care and Research, Boston University Medical Center, Boston, Massachusetts.

required HCV RNA testing for diagnosis.[8] Thus, additional investigation is needed to understand the mechanisms underlying the association between HIV/HCV coinfection and renal disease.

Individuals with HIV/HCV coinfection demonstrate an accelerated course of liver disease progression to cirrhosis,[9] and experience a high degree of liver-related morbidity and mortality.[10] AKI is a frequent complication of decompensated cirrhosis in the general population, and in Franceschini's study,[4] liver failure accounted for 18% of AKI events among HIV/HCV-coinfected subjects. A subsequent study in the United Kingdom demonstrated that among 20 HIV/HCV-coinfected patients with AKI, 7 had advanced cirrhosis while 5 experienced infectious complications of injection drug use (IDU).[7]

Although decompensated liver cirrhosis and infectious complications of IDU underlie some of the etiologies by which HIV/HCV coinfected patients develop AKI, additional mechanisms remain unexplored. HCV has independently been associated with immune complex-mediated renal injury, but epidemiologic studies of the association between HCV monoinfection and renal disease have been inconclusive. Several studies have found increased associations between HCV infection and albuminuria, CKD, and end-stage renal disease (ESRD),[11–13] while others have found no association.[14] Since most studies of HIV/HCV infection and renal disease have relied solely on HCV antibody data, the impact of various HCV-specific factors such as viral load and genotype have not been well studied. In addition, despite the high prevalence of substance use among coinfected patients, the associations between various substances of abuse and AKI have not been elucidated.

In this study, we sought to determine the incidence of AKI in HIV/HCV-coinfected versus HCV-monoinfected subjects, and to examine the association between HCV-specific factors and exposure to various substances of abuse, and the subsequent development of AKI.

Methods

Study design/subject assembly

Subjects with HIV/HCV coinfection and HCV monoinfection were enrolled in a prospective cohort study of the natural history of liver disease progression. The present analysis covers the time period August 15, 2000 to December 31, 2007. Subjects were included if they had at least two creatinine (Cr) measures obtained a minimum of 3 months apart, and were excluded if they had less than 6 months of follow-up. Subjects were categorized as HCV monoinfected or HIV/HCV coinfected based on their status at study entry.

Data collection

The study database contains information collected prospectively through semiannual patient surveys and annual electronic medical chart review. Data are available on self-reported illicit drug and alcohol use, clinical laboratory data, hospitalizations, liver events, and deaths.

Main exposures of interest Subjects reporting any quantity of heroin or cocaine use (through any route) within 6 months of interview were defined as having a positive exposure. Hazardous alcohol use was defined as 5 or more drinks at least once per month. Nonhazardous alcohol use was defined as 5 or more drinks less than once per month or 1–4 drinks at any time. Nondrinkers were defined as having no alcohol use within 6 months of interview. Liver events, signifying the presence of decompensated cirrhosis, included: hepatic encephalopathy, ascites, spontaneous bacterial peritonitis, gastrointestinal variceal bleed, and hepatocellular carcinoma. HCV genotypes were grouped into categories 1 versus non-1 as the majority of subjects were infected with genotype 1. Subjects were defined as having a high versus low HCV VL, dichotomized at 750,000 copies per milliliter, which was the mean VL of the cohort. Among HIV/HCV-coinfected subjects, HIV VL levels were defined as undetectable (<50 copies per milliliter), intermediate (50–4000 copies per milliliter), and high (>4000 copies per milliliter) based upon previous findings of an association between renal failure and HIV VL greater than 4000 copies per milliliter.[15]

Case ascertainment Electronic chart reviews were performed to identify episodes of incident AKI. For subjects with baseline Cr 1 mg/dL or less, AKI was defined as an absolute rise in Cr to more than 1.5 mg/dL or a relative rise of 0.5 mg/dL. For subjects with baseline Cr 1–2 mg/dL, AKI was defined as a rise in Cr of 0.5 mg/dL or more. For those with baseline Cr 2–4.9 mg/ dL, AKI was defined as a rise in Cr of 1.0 mg/dL or more. For those with baseline Cr greater than 5 mg/dL, AKI was defined as a rise in Cr of 1.5 mg/dL or more, as described previously.[1,4] When a rise in serum Cr occurred, available clinical data were reviewed to determine the probable cause. Events were categorized as prerenal, intrinsic renal, obstructive, or unknown. An event was attributed to prerenal azotemia when there was a clinical history of volume depletion and/or hypotension and renal function improved with hydration. Events were attributed to instrinsic renal disease from ischemic or toxic injury, based on available clinical data. Events were attributed to obstructive renal disease if clinical or radiographic evidence of nephrolithiasis or other forms of obstructive disease were identified. If data were insufficient, cause was characterized as unknown. Cases were reviewed by a nephrologist (M.H.) to ensure that criteria were met and events consistently characterized.

Statistical analysis

Data were analyzed using SAS version 9.1 (SAS Institute, Cary, NC). AKI incidence rates (IR) were calculated by dividing number of events by person time at risk. IRs were calculated for first AKI events, but not for subsequent events. $\chi 2$ and Fisher exact tests were utilized to compare baseline characteristics between HIV/HCV-coinfected and HCV-monoinfected subjects. Kaplan-Meier estimates of time to AKI were produced and stratified by infection status. For subjects who died or were lost to follow-up, data was censored at the time of study drop out. Data was censored on December 31, 2007 for all subjects who reached this date without developing AKI. Cox proportional hazards were utilized to determine risk factors for AKI. Variables significant in univariate analysis, as well as potential confounders, were included in multivariable analysis. Due to concern for collinearity between heroin, cocaine, and alcohol use, adjusted models were run separately for each of these variables prior to their inclusion

into a final model. Certain variables including liver event history and heroin, cocaine, or hazardous alcohol use within the past 6 months, were also evaluated as time-varying covariates at 6-month intervals. All tests were two tailed with a significance level of 0.05.

Results

Two hundred sixteen HIV/HCV-coinfected and 151 HCV-monoinfected subjects were included in the analysis. Other than higher rates of cocaine use among coinfected subjects, no significant differences in baseline characteristics were found. Table 1 summarizes baseline characteristics. A substantial percentage of HIV/HCV coinfected and HCV monoinfected subjects demonstrated use of cocaine, heroin, or hazardous alcohol within 6 months of study entry.

AKI incidence

Among HIV/HCV-coinfected subjects, 75 first AKI events occurred over 858 person-years, with an IR of 8.74 per 100 person-years. Among HCV-monoinfected subjects, 25 first AKI events occurred over 708 person-years, with an IR of 3.53 per 100 person-years. Kaplan Meier estimates of time to AKI by infection status are displayed in Fig. 1.

Etiology of AKI

Of the 100 AKI events, 45% were due to prerenal causes, 31% were due to intrinsic renal causes, and one case was due to obstructive nephrolithiasis from indinavir use. In 23% of cases, the mechanisms of AKI could not be determined. Of the 31 cases of AKI due to intrinsic renal etiologies, three cases each were due to acute interstitial nephritis, acute tubular necrosis, and rhabdomyolysis. Two cases each were attributed to renal crystaluria secondary to high-dose acyclovir, amphotericin B treatment for cryptococcal meningitis, and staphylococcal sepsis. One case each was attributed to complications of chemotherapy and pentamidine treatment for PCP pneumonia. Only one case was attributed to HIV-associated nephropathy. The etiology of intrinsic AKI could not be determined for 13 cases.

Table 1. Baseline Characteristics by Infection Status

Characteristic	*HIV/HCV* n = *216 (%)*	*HCV* n = *151 (%)*	p *Value*
Male gender	143 (66)	86 (57)	0.15
Age $\leq$ 45 years	119 (55)	74 (49)	0.25
Race/ethnicity			
Black/not Hispanic	110 (51)	70 (46)	0.18
White	49 (23)	49 (32)	
Latino/Hispanic	52 (24)	28 (19)	
Hypertension[a]	82 (38)	63 (42)	0.49
Diabetes[a]	40 (19)	38 (25)	0.13
Liver event	27 (13)	14 (9)	0.33
HCV genotype			
1	154 (72)	118 (78)	0.18
2–4	39 (18)	26 (17)	
HCV viral load $\geq$750,000 copies/mL	106 (49)	60 (40)	0.11
Cocaine use[b]	56 (26)	26 (17)	0.04
Heroin use[b]	48 (22)	35 (23)	0.81
Hazardous alcohol[b]	54 (25)	41 (27)	0.78

[a]Indicates presence of characteristic at any point during study period.
[b]Indicates use within 6 months of study entry.
HCV, hepatitis C virus.

Univariate analysis

Table 2 summarizes results of univariate analysis of baseline risk factors for AKI. Significant predictors of AKI included black race, hypertension, HIV/HCV coinfection, liver event, cocaine use, and hazardous drinking within 6 months of study entry.

Multivariable analysis

Successive multivariable Cox proportional hazards regression models showed the following variables to be significant independent risk factors for AKI: cocaine use within 6 months of study entry (HR 2.06 [95% CI 1.15, 3.71]), history of liver event ever (HR 6.64 [95% CI 3.81, 11.6]), and HIV/HCV coinfection (HR 2.19 [95% CI 1.33, 3.62]). HCV-specific factors, including genotype and HCV VL, did not show significant associations in univariate or adjusted analyses. Results of multivariable analyses are summarized in Table 3. While hazardous alcohol use is associated with AKI in univariate analysis, the association loses statistical significance in the adjusted analysis. Heroin use appears to be significantly associated with AKI in an adjusted model, but loses statistical significance when cocaine use is subsequently added into the model.

Certain variables, including liver event, cocaine, heroin, and hazardous alcohol use, were also evaluated as time-varying covariates over 6-month intervals throughout the study period. In an adjusted model, liver event (HR 4.70 [95% CI 2.85, 7.75]) and cocaine use (HR 1.92 [1.10, 3.36]) within the past 6 months remained significantly associated with AKI over time.

Subgroup analysis in HIV/HCV-coinfected subjects

Among subjects with HIV/HCV coinfection, additional analysis of risk factors for AKI was performed with adjustment for CD4 cell count and HIV VL. In univariate analysis, baseline CD4 cell count less than 200/mm^3 was significantly associated with AKI (HR 1.90 [95% CI 1.17, 3.31]) compared to CD4 cell count greater than 350/ mm^3. HIV VL greater than 4000 copies per milliliter was also associated with AKI (HR 1.53 [95% CI 0.90, 2.60]) compared to HIV VL less than 50 copies per milliliter, although not statistically significant. Results of multivariable analyses among coinfected subjects, adjusting first for CD4 cell count and then for HIV viral load, are summarized in Tables 4 and 5. Cocaine use within 6 months of study entry and history of a liver event remained significantly associated with AKI, even after adjustment for CD4 cell count and HIV VL.

HAART and AKI

Antiretroviral regimens (ARVs) at the time of AKI were examined for the 75 HIV/HCV-coinfected subjects. Forty percent of subjects were not on ARVs at the time of AKI. Of the 45 subjects on ARVs, 38% were on tenofovir-containing regimens, 7% were on indinavir-containing regimens, and

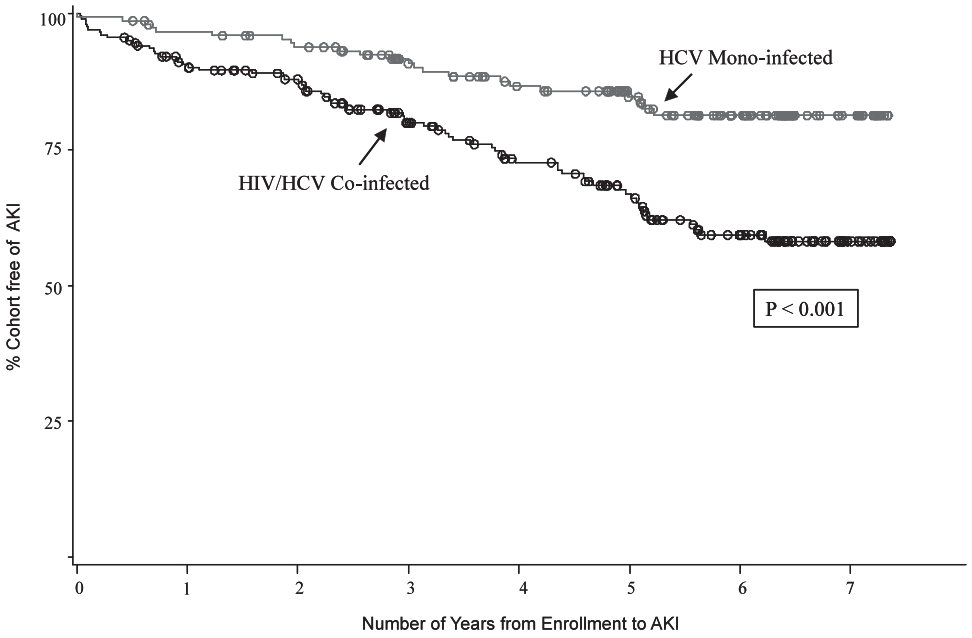


FIG. 1. Kaplan Meier curves of time to incident acute kidney injury (AKI) stratified by infection status. Mean follow up time was 5.15 years.

55% were on other ARV regimens. Collection of tenofovir data for the cohort did not begin until 2005 and was inconsistent. Upon query of medication logs, we determined that 74 (34%) of HIV/HCV-coinfected subjects reported ever taking tenofovir. In univariate analysis, the RR of AKI for HIV/HCV coinfected subjects reporting tenofovir use versus no use was 0.70 (95% CI [0.44–1.13]); *p* value 0.137.

AKI associated morbidity and mortality

The majority of AKI events were associated with either an emergency department visit or hospitalization, including 68% of episodes among HCV-monoinfected subjects and 89% of episodes among HIV/HCV coinfected subjects. While no AKI events among HCV-monoinfected subjects resulted in need for hemodialysis, six (8%) events among HIV/HCV-coinfected subjects led to a transient need for hemodialysis. AKI resulted in death in one (0.4%) HCV-monoinfected subject and six (8%) HIV/HCV coinfected subjects. Results are displayed in Fig. 2.

Table 2. Univariate Analysis of Risk Factors for Acute Kidney Injury

Variable	*HR (95% CI)*	p *Value*
Male gender	1.03 (0.69, 1.53)	0.89
Age $\leq$ 45	0.72 (0.49, 1.07)	0.11
Black race[a]	1.60 (1.07, 2.39)	0.02
Hypertension	1.80 (1.22, 2.67)	<0.001
Diabetes	1.54 (0.99, 2.41)	0.06
Liver event	4.68 (2.96, 7.41)	<0.001
HIV/HCV infection[b]	2.48 (1.57, 3.89)	<0.001
HCV genotype 2–4[c]	0.93 (0.52, 1.65)	0.80
HCV VL $\geq$ 750,000 copies/mL	1.05 (0.69, 1.60)	0.82
Cocaine use	1.72 (1.13, 2.61)	0.01
Heroin use	1.18 (0.75, 1.87)	0.43
Hazardous alcohol use[d]	1.70 (1.08, 2.67)	0.02

Reference groups are as follows:
[a]Non-black race/ethnicity (including white, Latino/Hispanic, Asian, and other).
[b]HCV monoinfection.
[c]HCV genotype 1.
[d]Moderate alcohol use or no alcohol use.
HR, hazard ratio; CI, confidence interval; HCV, hepatitis C virus; VL, viral load.

Discussion

In this analysis, AKI incidence was 2.2 times higher in subjects with HIV/HCV coinfection than in HCV monoinfection. In comparison to Franceschini's study,[5] in which subjects with HIV/HCV coinfection and HIV monoinfection had a combined incidence of first AKI event of 4.3 per 100 person-years, AKI incidence was twofold higher (8.74 per 100 person-years) among HIV/HCV-coinfected subjects in our

Table 3. Multivariable Analysis of Risk Factors for Acute Kidney Injury

Model	*Variable*	*HR (95% CI)*	p *Value*
1	Cocaine use	1.93 (1.23, 3.02)	0.004
2	Heroin use	1.62 (1.01, 2.60)	0.045
3	Hazardous alcohol	1.49 (0.94, 2.36)	0.092
4	Cocaine use	1.76 (1.07, 2.87)	0.025
	Heroin use	1.28 (0.76, 2.14)	0.352
5	Cocaine use	2.06 (1.15, 3.71)	0.016
	Heroin use	0.97 (0.53, 1.77)	0.910
	Hazardous alcohol	1.33 (0.76, 2.32)	0.314
	HCV genotype 2–4	0.84 (0.44, 1.60)	0.595
	HCV viral load $\geq$ 750,000 copies/mL	0.96 (0.60, 1.54)	0.877
	Liver event	6.64 (3.81, 11.6)	<0.001
	HIV/HCV infection	2.19 (1.33, 3.62)	0.002

All Cox proportional hazards regression models adjusted for age, gender, race/ethnicity, diabetes, hypertension, liver event ever, and infection status.
HR, hazard ratio; HCV, hepatitis C virus.

Table 4. Multivariable Analysis of Risk Factors for Acute Kidney Injury, Adjusted for CD4 Cell Count

Variable	*HR (95% CI)*	p *Value*
CD4 cell count		
>350/mm^3 (reference group)	—	—
201–350/mm^3	0.59 (0.31, 1.10)	0.0960
<200/mm^3	1.76 (1.02, 3.05)	0.0424
Liver event	5.63 (3.11, 10.21)	<0.0001
Cocaine use	2.32 (1.39, 3.88)	0.0013

Cox proportional hazard regression model is adjusted for age, race/ethnicity, diabetes, and hypertension.
HR, hazard ratio; CI, confidence interval.

current analysis. After adjusting for the elevated risk of AKI attributed to HIV coinfection, additional significant risk factors for AKI included decompensated liver cirrhosis and cocaine use. These risk factors remained significant among HIV/HCV-coinfected subjects, even after adjustment for CD4 cell count and HIV VL.

In contrast to findings that higher HIV VL is associated with increased risk of AKI, we did not find an association between HCV VL and AKI incidence in our analysis. While several studies have demonstrated that HIV infection may be linked to renal disease through direct viral infection of renal parenchyma,[16] a similar mechanism has not been demonstrated for HCV infection. HCV has been linked to several types of glomerular lesions including membranoproliferative glomerulonephritis (MPGN) and membranous glomerulopathy. These pathologic lesions appear to result from deposition of circulating immune complexes that contain HCV and anti-HCV antibody. Although a few reports of immunohistochemical localization of HCV antigen in renal tissue have been published, results have not been consistently demonstrated.[16,17] Although in our study, higher HCV VL and genotype were not associated with AKI, these findings add important information to currently available data as other epidemiologic studies examining the relationship between HCV and renal disease have not included HCV VL or genotype data.[11–14]

In addition to HIV/HCV coinfection and decompensated liver cirrhosis, cocaine but not heroin use was strongly associated with AKI. Indeed, the entity previously known as heroin nephropathy has greatly decreased in recent years and likely represented primary focal and segmental glomerulosclerosis (FSGS) rather than a consequence of heroin use.[18] FSGS is now the most common cause of nephrosis in African American individuals in the United States. In contrast, cocaine clearly causes deleterious effects on the kidney and can lead to AKI both from rhabdomyolysis and malignant hypertension.[19,20] This study further supports the existence of an association between cocaine use and AKI among persons with HIV/HCV coinfection.

Table 5. Multivariable Analysis of Risk Factors for Acute Kidney Injury, Adjusted for HIV Viral Load

Variable	*HR (95% CI)*	p *Value*
HIV viral load		
<50 copies/mL (reference group)	—	—
51–4000 copies/mL	0.97 (0.53, 1.79)	0.9321
>4000 copies/mL	1.72 (0.98, 3.03)	0.0583
Liver event	5.85 (3.23, 10.61)	<0.0001
Cocaine use	2.19 (1.31, 3.67)	0.0028

Cox proportional hazards regression model is adjusted for age, race/ethnicity, diabetes, and hypertension.
HR, hazard ratio; CI, confidence interval.

Among the HIV/HCV-coinfected subjects, there was only a single case of HIV-associated nephropathy identified. No cases of other diagnoses attributable to HIV infection, such as HIV immune complex kidney disease, IgA nephropathy or thrombotic thrombocytopenic purpura (TTP), were found. The bulk of cases were related to prerenal azotemia or drug toxicity. In multivariable analyses, lower CD4 cell count and higher HIV VL were both associated with AKI, although only CD4 cell count less than 200 was statistically significant. Lower CD4 count may have been associated with AKI through higher rates of nephrotoxic drug use for opportunistic infections and ARV toxicity. Indinavir-associated renal dysfunction is well documented and there is also growing evidence of an association between tenofovir use and loss of renal function.[22–27] In a recent meta-analysis that included 17 studies to assess the renal safety of tenofovir in HIV-infected patients, a statistically significant greater loss in creatinine clearance was observed among tenofovir recipients compared to control subjects, as well as a significantly greater risk of acute renal failure among tenofovir recipients.[22] Of subjects on ARVs at the time of AKI, 45% were receiving indinavir or tenofovir-containing regimens, reflecting evolving drug availability and prescribing practices during the study period. However, sufficient data were not available to examine the independent association between tenofovir use and AKI. The interrelationships between CD4 cell count, HIV VL, exposure to ARVs, and the subsequent development of AKI warrant further investigation among HIV/HCV-coinfected subjects.

There were several limitations to this analysis. Serial HIV antibody testing was not performed as part of the research protocol after study entry, and could have resulted in misclassification of some dually infected subjects as having HCV monoinfection. Such a misclassification may have dampened the true impact of HIV coinfection on the incidence of AKI among subjects with HCV infection. HCV VL and genotype data was missing for 12% of the cohort and may have diminished the power to detect a true association between HCV VL and AKI. Furthermore, kidney biopsies were not performed in this study to look for direct HCV viral effects on renal tissue. Data on tenofovir use were not consistently collected for the cohort and the data presented likely underestimates the true proportion, of individuals that were taking tenofovir. Due to concerns with the quality of the tenofovir data and associated biases, we did not adjust for tenofovir use in the final multivariable analyses. Thus, tenofovir use may have been a potential confounder of the association between variables of interest and AKI among coinfected subjects. As no standard definition for AKI exists, a traditional definition of AKI, which has been shown to affect mortality in hospitalized patients without HIV or HCV, was utilized.[4] Recently, smaller changes in serum Cr have been found to be of clinical import, such that a rise in Cr of 0.5 mg/dL is utilized in the RIFLE criteria (Risk Injury Failure Loss End Stage) and a rise of only 0.3 mg/dL is utilized in the AKIN criteria (Acute Kidney

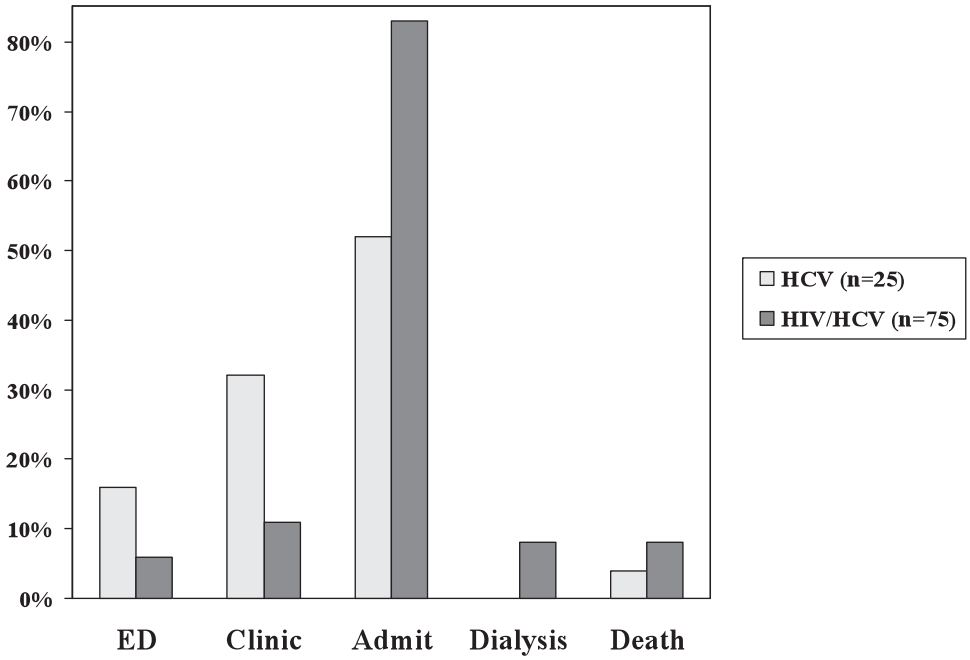


FIG. 2. Percentage of acute kidney injury (AKI) episodes associated with emergency department visit, clinic visit, hospital admission, need for transient or permanent dialysis, and death. Results are broken down by infection status.

Injury Network).[28] In our study, we chose to utilize the more conservative definition of AKI to capture the most serious events.

The findings of the association between AKI and decompensated liver cirrhosis, HIV coinfection, and cocaine use in this study, highlight the importance of controlling HIV infection in coinfected individuals and preventing liver disease progression and cocaine use in persons with both HIV/HCV coinfection and HCV monoinfection. While the relationship between AKI and other substances of abuse, including heroin and hazardous alcohol, was not statistically significant after adjusting for cocaine use, intermediate multivariable models did demonstrate a positive association between these agents and AKI. Thus, targeted interventions to reduce heroin and alcohol use may further reduce the burden of renal disease among these individuals. Further studies are required to determine the association between HCV VL and AKI, and whether lowering HCV VL may impact AKI incidence. By addressing the identified modifiable risk factors, a reduction in AKI-associated morbidity and mortality, including rates of hospitalization, transient hemodialysis and death, may potentially be achieved in this population.

Acknowledgments

Funding for the CHARM cohort at Boston Medical Center is supplied by National Institute on Drug Abuse Grant 5R01DA019841-09. Additional funding provided to S.G. through National Institute of Health Institutional Research Training Grants 5T32AI007433-18 and 5T32AI007438-17. The authors thank the study participants, investigators, study coordinators, and research interviewers at Boston Medical Center, and the data coordinating center at Boston University School of Public Health, for making this study possible.

Author Disclosure Statement

No competing financial interests exist.

References

1. Gupta SK, Eustace JA, Winston JA, et al. Guidelines for the management of chronic kidney diseases in HIV- infected patients: Recommendations of the HIV Medicine Association of the Infectious Diseases Society of America. Clin Infect Dis 2005;40:1559–1585.
2. Fine DM, Atta MG. Kidney disease in the HIV-infected paient. AIDS Patient Care STDs 2007;21:813–824.
3. Wyatt CM, Arons RR, Klotman PE, Klotman ME. Acute renal failure in hospitalized patients with HIV: Rsk factors and impact on in-hospital mortality. AIDS 2006;20:561–5605.
4. Franceschini N, Napravnik S, Eron JJ Jr, Szczech LA, Finn WF. Incidence and etiology of acute renal failure among ambulatory HIV- infected patients. Kidney Int 2005;67:1526–1531.
5. Franceschini N, Napravnik S, Finn WF, Szczech LA, Eron JJ Jr. Immunosupression, hepatitis C infection, and acute renal failure in HIV- infected patients. J Acquir Immune Defic Syndr 2006;42:368–372.
6. Lopes JA, Fernandes J, Jorge S, Neves J, Antunes F, Prata MM. Acute renal failure in critically ill HIV- infected patients. Crit Care 2007;11:404.
7. Roe J, Campbell LJ, Ibrahim F, Hendry BM, Post FA. HIV Care and Incidence of Acute Renal Failure. Clin Infect Dis 2008;47:242–249.
8. Wyatt CM, Malvestutto C, Coca SG, Klotman PE, Parikh CR. The impact of hepatitis C virus coinfection on HIV-related kidney disease: A systematic review and meta-analysis. AIDS 2008;22:1799–1807.
9. Graham CS, Baden LR, Yu E, et al. Influence of human immunodeficiency virus infection on the course of hepatitis C virus infection: A meta-analysis. Clin Infect Dis 2001;33:562–569.
10. Weber R, Sabin CA, Friis- Moller N et al. Liver-related deaths in persons infected with the human immunodeficiency virus: The D:A:D study. Arch Intern Med 2006;166:1632–1641.
11. Tsui JI, Vittinghoff E, Shlipak MG, O'Hare AM. Relationship between hepatitis C and chronic kidney disease: Results

from Third NHANES Survey. J Am Soc Nephrol 2006; 17:1168–1174.
12. Dalrymple LS, Koepsell T, Sampson J, et al. Hepatitis C Virus Infection and the Prevalence of Renal Insufficiency. Clin J Am Soc Nephrol 2007;2:715–721.
13. Tsui JI, Vittinghoff E, Shlipak MG, et al. Association of hepatitis C seropositivity with increased risk for developing end-stage renal disease. Arch Intern Med 2007;167: 1271–1276.
14. Moe SM, Pampalone AJ, Ofner S, Rosenman MD, Teal E, Hui SL. Association of hepatitis C virus infection with prevalence and development of kidney disease. Am J Kidney Dis 2008;51:885–892.
15. Szczech LA, Gange SJ, van der Horst C, et al. Predictors of proteinuria and renal failure among women with HIV infection. Kidney Int 2002;61:195–202.
16. Alpers CE, Kowalewska J. Emerging paradigms in the renal pathology of viral diseases. Clin J Am Soc Nephrol 2007;2:S6–S12.
17. Barsoum RS. Hepatitis C virus: From entry to renal injury-facts and potentials. Nephrol Dial Transplant 2007;22:1840–1848.
18. Jaffe JA, Kimmel PL. Chronic nephropathies of cocaine and heroin abuse: A critical review. Clin J Am Soc Nephrol 2006;1:655–667.
19. Fine DF, Garg N, Haas M, et al. Cocaine use and hypertensive renal changes in HIV-infected individuals. Clin J Am Soc Nephrol 2007;2:1125–1130.
20. Norris KC, Thornhill-Joynes M, Robinson C, et al. Cocaine use, hypertension, and end-stage renal disease. Am J Kidney Dis 2001;38:523–528.
21. Glauser J, Queen JR. An overview of non-cardiac cocaine toxicity. J Emerg Med 2007;32:181–186.
22. Cooper RD, Wiebe N, Smith, N, Keiser P, Naicker S, Tonelli M. Systematic review and meta-analysis: Renal safety of tenofovir disoproxil fumarate in HIV-infected patients. Clin Infect Dis 2010;51:496–505.
23. Gallant JE, Staszwski S, Pozniak AI, et al. Efficacy and safety of tenofovir DF vs stavudine in combination therapy in antiretroviral-naïve patients: A 3-year randomized trial. JAMA 2004;292:191–201.
24. Young B, Buchacz K, Moorman A, Wood K, Brooks JT; the HIV Outpatient Study (HOPS) Investigators. Renal function in patients with preexisting renal disease receiving tenofovir-containing highly active antiretroviral therapy in the HIV outpatient study. AIDS Patient Care STDs 2009;23:589–592.
25. Crum-Cianflone N, Ganesan A, Teneza-Mora N et al. Prevalence and factors associated with renal dysfunction among HIV-infected patients. AIDS Patient Care STDs 2010;24:353–360.
26. Verhelst D, Monge M, Meynard JL, et al. Fanconi syndrome and renal failure indcued by tenofovir: A first case report. Am J Kidney Dis 2002;40:1331–1333.
27. Szczech LA. Renal dysfunction and tenofovir toxicity in HIV infected patients. Top HIV Med 2008;16:122–126.
28. Mehta RL, Kellum JA, Shah SV, et al. Acute kidney injury network: Report of an initiative to improve outcomes in acute kidney injury. Crit Care 2007;11:R31.

Address correspondence to:
Shikha Garg
1600 Clifton Road
Mailstop A-32
Atlanta, GA 30333

E-mail: shikhagarg75@yahoo.com

Update in Addiction Medicine for the Generalist

Adam J. Gordon, MD, MPH[1], Hillary V. Kunins, MD, MPH, MS[2], Darius A. Rastegar, MD[3], Jeanette M. Tetrault, MD[4], and Alexander Y. Walley, MD, MSc[5]

[1]Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System and University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; [2]Albert Einstein College of Medicine and Montefiore Medical Center, New York, NY, USA; [3]Johns Hopkins School of Medicine, Baltimore, MD, USA; [4]Yale University School of Medicine, New Haven, CT, USA; [5]Boston University School of Medicine, Boston, MA, USA.

KEY WORDS: review; substance-related disorders; primary health care; primary care; alcoholism and addictive behavior; drug abuse; smoking cessation.
J Gen Intern Med 26(1):77–82
DOI: 10.1007/s11606-010-1461-3

INTRODUCTION

Generalist clinicians routinely care for patients who misuse or are dependent on alcohol, nicotine, and other drugs of abuse.[1,2] These problems contribute to significant morbidity, health care utilization, cost, and preventable death.[3,4] The aim of this update is to identify and examine recent advances in addiction medicine that have practice implications for generalist physicians and their patients. To accomplish this, we independently selected articles in the field of addiction medicine, summarized and critically appraised, and examined the articles in the context of their implications for generalist practice using methodology we used in prior updates.[5,6] During an initial review, we identified articles through an electronic MedLine search (limited to human studies and in English) using search terms for alcohol, nicotine, and other drugs of abuse from January 2008 through January 2010. From the citations, the authors selected articles for more intensive review. After this initial review, we searched for other literature in web-based or journal resources (e.g., www.aodhealth.org, ACP Journal Club, table of contents of relevant journals). All authors then agreed collectively on the important articles regarding addiction medicine that have implications for practice for generalist clinicians.

PRESCRIPTION DRUG ABUSE

In treating chronic pain, physicians must balance the pain-relieving benefits of opioids with the risks of overdose and triggering addiction. Efforts to improve pain treatment since the 1990s have led to increases in opioid prescriptions. At the same time, there have been substantial increases in misuse and diversion of prescription opioids, opioid-related emergency department visits, and fatal opioid overdoses.

Received May 28, 2010
Revised June 30, 2010
Accepted July 6, 2010
Published online August 10, 2010

What Factors Increase the Overdose Risk of Prescribed Opioids?

Dunn KM and colleagues. Opioid prescriptions for chronic pain and overdose: a cohort study. Annals of Internal Medicine.2010;152(2):85–92.[7]

Dunn and colleagues sought to estimate rates of fatal and non-fatal opioid overdose and determine whether these rates vary by prescribed opioid dose among patients receiving medically prescribed, long-term opioid therapy.[7] By linking pharmacy, electronic medical and state mortality records, investigators evaluated outcomes among 9,940 persons in a health maintenance organization who received three or more opioid prescriptions within 90 days for chronic non-cancer pain between 1997 and 2005. They estimated non-fatal and fatal overdose risk as a function of average daily opioid dose (morphine equivalents) received at the time of overdose. Over a mean follow-up time of 42 months, they identified 51 patients with opioid-related overdoses, 6 of whom died (mean follow-up time of 42 months). The rate of any opioid overdose was 0.15 per 100 person-years, and the rate of overdose mortality was 0.02 per 100 person-years. Overdose rates were found to be dose-related; compared with patients receiving 1 to 20 mg/day of opioids, patients receiving 100 mg/day or more had an 8.9 fold (95% CI: 4.0–19.7) increase in overdose risk. Overdose rates were also increased by two to three fold in patients who had a history of a substance abuse diagnosis, depression, or were receiving a concomitant sedative-hypnotic prescription.

Are Increases in Overdose Deaths Related to the Diversion of Prescription Drugs?

Hall AJ and colleagues. Patterns of abuse among unintentional pharmaceutical overdose fatalities. Journal of the American Medical Association.2008;300(22):2613–20.[8]

Hall and colleagues sought to evaluate the risk characteristics of persons dying of unintentional prescription drug overdose in West Virginia in 2006.[8] They linked data from the state's medical examiner database, prescription drug monitor-

ing program, and opiate treatment programs to describe the type of substance use and concomitant behaviors. Among 295 people who died from an unintentional prescription drug poisoning, 78% had a history of substance abuse, 63% had taken a prescription drug not prescribed, 21% had 5 or more different prescribers over 12 months, 17% had a previous overdose, 16% had cocaine, heroin or methamphetamine also present, and 4% were enrolled in a methadone maintenance program. Prevalence of prescription drugs not prescribed was greatest among decedents aged 18 through 24 years and decreased across each successive age group. More than one substance was detected in 79% of deaths. The most common substance was opioids (93% of deaths); of these, only 44% had ever been prescribed these drugs. Methadone was the most common opioid (40%), followed by hydrocodone (23%) and oxycodone (21%). However, only 32% of decedents with methadone present at death had a prescription for it, whereas 85% with hydrocodone and 61% with oxycodone had prescriptions for hydrocodone and oxycodone, respectively. The authors concluded that the majority of prescription drug overdose deaths in West Virginia in 2006 were associated with nonmedical use and diversion of prescription drugs, primarily opioid analgesics.

What is the Arrhythmia Risk from Methadone?

Anchersen K and colleagues. Prevalence and clinical relevance of corrected QT interval prolongation during methadone and buprenorphine treatment: a mortality assessment study. Addiction.2009;104:993–999.[9]

Studies conducted by Hall and others have documented an increasing and disproportionate prevalence of methadone-related deaths, which has led to increasing focus on QT prolongation in methadone patients and the risk of torsades de pointes.[8,10] In this setting, recent expert panel recommendations for universal electrocardiography (ECG) screening and regular monitoring for corrected QT (QTc) prolongation for patients prescribed methadone[11] have been challenged as reaching beyond the evidence.[12] Anchersen determined the maximum mortality rate potentially attributable to QTc prolongation by linking the Norwegian Opioid Maintenance Treatment (OMT) registry and the national death certificate register.[9] They found 90 deaths occurring among 2,382 patients between 1997 and 2003, a rate of 1.3/100 patient-years. After review of each of these case records, four deaths were identified in which QTc prolongation could not be excluded as the cause of death. Thus, at most, 4% of methadone deaths could be potentially attributable to arrhythmias, resulting in a maximum mortality rate of 0.06 per 100 patient-years.

IMPLICATIONS FOR PRACTICE

Among patients treated with prescription opioids for chronic pain, having a substance use disorder, concomitant prescriptions for sedative-hypnotics, and higher doses of opioids increase the risk of overdose. However, most overdose deaths from prescription drugs involve diverted medications, particularly those involving methadone. Overdose victims commonly have a substance abuse history, mix multiple substances, and seek prescriptions from multiple prescribers. Methadone is disproportionately involved in overdoses, yet this is not explained by overdoses among patients receiving methadone maintenance for opioid dependence. Although methadone prolongs the QTc interval, other factors, such as prolonged and variable metabolism and mixing methadone with other sedating substances, likely explain the disproportionate number of deaths with methadone present compared to other opioids. Prescribers and patients should be educated about these overdose risk factors. Opioid prescribers should have a goal-directed approach, continuing or increasing opioid therapy only when there is demonstrated improvement in function or quality of life.[13] They should consider strategies to assess adherence and limit diversion, such as prescription monitoring programs, toxicology testing, and pill counts.

ADDICTION SCREENING AND BRIEF INTERVENTIONS

Cost Effectiveness of Screening for Addictions

Solberg LI and colleagues. Primary care intervention to reduce alcohol misuse: ranking its health impact and cost-effectiveness. American Journal of Preventive Medicine.2008;34(2):143–152.[14]

The US Preventive Services Tasks Force (USPSTF) recommends screening and behavioral counseling interventions in primary care to reduce alcohol misuse. Screening, Brief Intervention, and Referrals and/to Treatment (SBIRT) is a strategy that has been tested in emergency room and primary care settings with proven efficacy, yet is incorporated into practice less than 9% of the time.[15] To measure the clinically preventable burden (CPB) and cost-effectiveness of screening and brief interventions (SBI) compared with other recommended preventive services, Solberg and colleagues conducted a systematic review of randomized controlled trials and cost-effectiveness studies.[14] CPB was calculated as the product of effectiveness and the alcohol-attributable fraction of mortality and morbidity (measured in quality-adjusted life years [QALYs]). Cost effectiveness was estimated from both the societal perspective and the health system perspective. Calculated CPB was 176,000 QALYs saved over the lifetime of a birth cohort of 4,000,000 individuals. Screening and brief counseling were cost-saving from the societal perspective and had a cost-effectiveness ratio of $1,755/QALY saved. SBI is one of the top five ranking preventive services, comparable to screening for colorectal cancer, hypertension, visual problems, and for influenza and pneumococcal vaccination.

Screening for Addictions in Primary Care

Smith PC and colleagues. Primary care validation of a single-question alcohol screening test. Journal General Internal Medicine.2009;24(7):783–8.[16]

Unhealthy alcohol use is under-diagnosed in primary care settings.[17] High performing screening instruments, such as the Alcohol Use Disorders Identification Test (AUDIT) and the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST), are lengthy and can be difficult to administer in primary care environments.[18–20] Several prior studies have

examined the performance characteristics of abbreviated and single-item alcohol screening questionnaires, but none have examined the single-item screening test recommended by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) in primary care settings.[21–24] Smith and colleagues validated, in an urban safety net primary care clinic, the performance of the single screening question, "How many times in the past year have you had X or more drinks in a day?" where X is 5 for men and 4 for women, and a response of ≥1 is considered positive.[16] They defined unhealthy alcohol use as the presence of an alcohol use disorder (either Alcohol Abuse or Dependence[25]) or risky consumption, as determined using a validated 30-day calendar method.[26,27] Among 286 patients, the single-question screen was 82% sensitive (95%, CI=73%–89%) and 79% specific (CI=73%–84%) for the detection of unhealthy alcohol use. It was more sensitive (88%, CI=73%–95 %) but less specific (67%, CI=61%–72%) for the detection of a current alcohol use disorder. Test characteristics were similar to that of a commonly used three-item screen (first three items of the AUDIT) and were affected very little by subject demographic characteristics. They concluded that the single screening question recommended by the NIAAA accurately identified unhealthy alcohol use in this sample of primary care patients. One potential limitation of this study was that the high prevalence of alcohol problems in this urban safety-net population (44%) may have been a marker for greater disease severity and/or selection bias, thus leading to greater sensitivity than would be observed in a lower risk population.

Application of SBIRT for Drug Use

Madras BK and colleagues. Screening, brief interventions, referral to treatment (SBIRT) for illicit drug and alcohol use at multiple healthcare sites: comparison at intake and 6 months later. Drug and Alcohol Dependence.2009;99:280–295.[28]

Components of SBIRT have been studied extensively for unhealthy alcohol use with evidence of efficacy, effectiveness, and cost-effectiveness.[29–33] However, research regarding the efficacy and effectiveness of SBIRT components for alcohol dependence and illicit drug use have been limited. In the last decade, an alcohol-focused SBIRT service program was initiated by the Substance Abuse and Mental Health Services Administration (SAMHSA) in a wide variety of medical settings (and diverse patient populations) in six states (http://www.sbirt.samhsa.gov). As part of this initiative, patients were screened and offered score-based progressive levels of intervention (brief intervention, brief treatment, referral to specialty treatment). Through a secondary analysis, Madras and colleagues examined illicit drug use at baseline and 6-month follow-up in a randomly selected sample of the nearly 60,000 patients who screened positive for drug use at baseline and received an SBIRT intervention.[28] Among those reporting baseline illicit drug use, rates of drug use at 6-month follow-up (4 of 6 sites) were 68% lower ($p<0.001$) and heavy alcohol use was 39% lower ($p<0.001$) than at baseline, with comparable findings across sites, gender, race/ethnic, and age subgroups. However, improvements in alcohol and illicit drug use from baseline were self-reported, assessed only in a sample of patients screening positive at baseline, and included subjects who all received an intervention, either brief intervention or more intensive treatment. Therefore, the true efficacy of brief intervention for illicit drug use needs further investigation.

IMPLICATIONS FOR PRACTICE

Generalist physicians must select among numerous recommended preventive medicine measures for their patients during brief visits. Solberg's study should encourage primary care physicians to prioritize SBIRT over other, less effective preventive interventions. Smith's study provides evidence that clinicians can screen for alcohol use disorders using the single-item alcohol screening question recommended by NIAAA. Moreover, the question detects alcohol use disorders, binge drinking, and risky drinking. Binge drinking is episodic, but deleterious to health, and risky drinking is a prevalent problem in primary care settings. Madras' evaluation of a large-scale SBIRT program implementation concluded that combined screening and intervention for alcohol and illicit drug use is feasible across a range of health care sites and diverse patient populations, though the true efficacy needs further investigation.

Applications of SBIRT-type interventions to populations other than at-risk drinkers have yielded mixed results, but remain an important area for ongoing investigation. A recent study of a counseling intervention among inpatients with prescription drug misuse showed a reduction in drug misuse at 3 months, but these improvements were no longer evident at 12 months.[34,35] A meta-analysis of 11 trials investigating brief intervention for hospitalized heavy drinkers had inconclusive results.[36] Finally, a study of brief intervention for dependent drinkers versus non-dependent drinkers enrolled in a clinical trial suggested that dependent drinkers decreased drinks per day similar to non-dependent drinkers 6 months following intervention.[37] Although promising, these studies do not provide definitive conclusions on the use of SBIRT for illicit drug use, prescription drug misuse, inpatient populations, or dependent drinkers. Further studies with controlled designs and standardized interventions are needed to assess the efficacy of SBIRT to these populations and settings.

Evidence for Interventions Associated with Co-Morbid Improvements in Health

Stewart SH and colleagues. Blood pressure reduction during treatment for alcohol dependence: results from the Combining Medications and Behavioral Interventions for Alcoholism (COMBINE) study. Addiction.2008;103:1622–28.[38]

Heavy alcohol consumption and alcohol use disorders are associated with a variety of health conditions, including hypertension. Stewart and colleagues evaluated blood pressure changes occurring during treatment for alcohol dependence among 1,383 subjects participating in the Combining Medications and Behavioral Interventions for Alcoholism (COMBINE) study, a large multi-center treatment study for alcohol dependence.[38,39] Over the 16-week treatment period, the authors assessed the relationship between percentage of drinking days (PDD) and systolic and diastolic blood pressure. Systolic blood pressure decreased by an average of 12 mmHg

and diastolic blood pressure decreased by an average of 8 mmHg; however, these reductions were only evident in people who were above the median blood pressure at baseline and occurred during the first month of treatment.

Nordback I and colleagues. The recurrence of acute alcohol-associated pancreatitis can be reduced: a randomized controlled trial. Gastroenterology. 2009;136(3):848–55.[40]

Alcohol-associated pancreatitis often recurs in patients who continue to consume alcohol. In this randomized clinical trial, Nordback and colleagues evaluated whether the recurrence of alcohol-associated acute pancreatitis can be reduced through a counseling intervention.[40] They examined 120 patients admitted to a university hospital for an initial episode of alcohol-associated acute pancreatitis and randomized them either to a 30-min, nurse-led pre-discharge intervention with repeated (6-month intervals) outpatient interventions or an initial, pre-discharge intervention only. They found that repeated interventions, each consisting of 30 min of counseling, appear to be better than a single intervention at hospital discharge in reducing the development of recurrent acute pancreatitis during a 2-year period.

IMPLICATIONS FOR PRACTICE

While alcohol can be deleterious to health, there has been scant but emerging literature indicating that interventions for alcohol may improve the effects of alcohol-related co-morbidities. Both the Stewart and Nordback studies suggest that a counseling intervention for alcohol use can have a positive impact on co-morbid health conditions. However, future research should focus on determining the appropriate intervention intensity (i.e., length and frequency of counseling) to produce improvements in outcomes of alcohol-related diseases commonly encountered in the clinical setting and whether alcohol counseling can improve other health outcomes.

ADDICTION PHARMACOTHERAPY

Pharmacotherapy for Smoking Cessation Treatment

Gunnell D and colleagues. Varenicline and suicidal behaviour: a cohort study based on data from the General Practice Research Database. British Medical Journal.2009;339:b3805.[41]

Pharmacotherapy is an important tool in helping smokers quit, and varenicline is an effective option.[42,43] However, reports of depression and suicidal thoughts have raised concerns about its safety.[44,45] Gunnell and colleagues investigated whether varenicline is associated with an increased risk of suicide and suicidal behavior when compared with other smoking pharmacotherapy. They evaluated fatal and non-fatal self-harm, suicidal thoughts, and depression among 80,660 patients in the UK who were prescribed a new course of a smoking cessation product; hazard ratios were adjusted for a number of factors, including current or previous psychiatric history. Those who received varenicline, when compared with nicotine replacement, had an adjusted hazard ratio (HR) for fatal or non-fatal self-harm of 1.17 with a 95% confidence interval of 0.59–2.32; the HR for suicidal thoughts was 1.43 (CI=0.53–3.85), and for start of antidepressant therapy, it was 0.88 (CI=0.77–1.00). There likewise was no significant difference in the adjusted hazard ratios associated with bupropion for any of these outcomes.

IMPLICATIONS FOR PRACTICE

While it is reasonable to warn patients of the potential for psychiatric side effects when prescribing varenicline, patients and providers can be reassured that the risk appears to be fairly low. Based on these data, approximately one out every 750 smokers who take varenicline for 3 months may experience an episode of self-harm. There is a possibility that all smoking pharmacotherapy, and perhaps smoking cessation itself, are associated with a modestly increased risk of self-harm and other psychiatric problems.

Office-Based Opioid Agonist Therapy

Walley AY and colleagues. Office-based management of opioid dependence with buprenorphine: clinical practices and barriers. J Gen Intern Med. 2008;23(9):1393–8.[46]

Office-based opioid-agonist therapy (OBOT) was made possible by the Drug Addiction Treatment Act of 2000, allowing physicians to prescribe approved medications (buprenorphine and buprenorphine/naloxone).[47] However, adoption by primary care physicians has been slow. Walley and colleagues assessed buprenorphine clinical practices and barriers in a survey mailed to all 225 office-based physicians in Massachusetts who were waivered to prescribe buprenorphine.[47] Prescribing physicians reported treating a median of ten patients; most non-prescribers (54%) reported they would prescribe if barriers were reduced. Factors associated with prescribing included being a primary care physician compared to a psychiatrist (AOR: 3.02; CI=1.48–6.18) and solo compared to group practice (AOR: 3.01; CI=1.23–7.35). On the other hand, reporting low patient demand (AOR: 0.043, CI=0.009–0.21) and insufficient institutional support (AOR: 0.37; CI=0.15–0.89) were associated with not prescribing.

Barry DT and colleagues. Integrating buprenorphine treatment into office-based practice: a qualitative study. Journal General Internal Medicine.2008;24(2):218–25.[48]

In a qualitative study, Barry and colleagues used semi-structured interviews of 23 office-based physicians in New England to identify physician, patient, and logistical factors that would either facilitate or serve as a barrier to OBOT.[48] Facilitators included promoting continuity of care, positive perceptions of buprenorphine, and viewing buprenorphine as a positive alternative to methadone. Physician barriers included competing activities, lack of interest, and lack of expertise in addiction treatment. Physicians' perceptions of patient-related barriers included concerns about confidentiality and cost, and low motivation for treatment. Perceived logistical barriers in-

cluded lack of remuneration for OBOT, limited ancillary support, time limitations, and a perceived low prevalence of opioid dependence.

Soeffing JM and colleagues. Buprenorphine maintenance treatment in a primary care setting: outcomes at 1 year. Journal Substance Abuse Treatment.2009;37(4):426–30.[49]

In clinical trials, buprenorphine has been shown to be efficacious at reducing illicit opioid use and improving other clinical outcomes.[50] The effectiveness has been supported by a number of recent observational studies in a variety of settings.[51,52] However, subjects in these studies were not treated in a manner typical of other chronic conditions in primary care practice, where the effectiveness of buprenorphine is less clear, particularly when onsite psychosocial services are not available. Soeffing and colleagues sought to investigate this question by assessing the 12-month outcomes of 255 patients given at least one prescription for buprenorphine in a primary care practice in Baltimore.[49] Patients were classified as "opioid-positive" or "opioid-negative" each month based on patient report, urine toxicology, and provider assessment. After 12 months, 145 (56.9%) patients remained in treatment, and the percentage who were opioid negative increased from 49% in the first month to 76% by month 12. These results are comparable to those reported in the landmark clinical trial of office-based buprenorphine, in which the percentage of opiate negative urine rose from 35% to 64% over a year.[53]

IMPLICATIONS FOR PRACTICE

The adoption of OBOT has the potential to expand treatment availability for opioid-dependent patients, particularly where methadone maintenance therapy is unavailable.[54] There are concerns and barriers that have limited the use of buprenorphine by physicians, even those who have undergone the required training and obtained a waiver.[55] Although not expressed in these studies, there is also a stigma associated with addiction and addiction treatment at a physician and institutional level that needs to be overcome.[56] Physicians who are interested in providing this treatment should be given institutional support and encouragement.

Observational studies support the effectiveness of office-based buprenorphine treatment for opioid dependence. There are a variety of approaches and treatment protocols, but providing this treatment in a setting where opioid dependence is incorporated into primary care and treated like other chronic illnesses appears to be effective. However, there are still unanswered questions about the optimal treatment approach, including intervals for visits, toxicology testing, and “dose” of counseling. It is likely that treatment needs to be individualized, and some patients will require a more intensive approach with closer monitoring and additional psychosocial support.

Acknowledgements: *We acknowledge the editorial assistance of Margaret Krumm in preparation of this manuscript. No financial support was received by the authors for the work represented in this manuscript. This work was presented in an invited oral presentation at the 2010 Annual meeting of the Society of General Internal Medicine. Minneapolis, Minneapolis, May 1, 2010.*

Conflict of Interest: *None disclosed*

Corresponding Author: *Adam J. Gordon, MD, MPH; Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System and University of Pittsburgh School of Medicine, 7180 Highland Drive (151-C-H), Pittsburgh, PA 15206, USA (e-mail: adam.gordon@va.gov).*

REFERENCES

1. **Robins LN, Helzer JE, Weissman MM, et al.** Lifetime prevalence of specific psychiatric disorders in three sites. Arch Gen Psychiatry. 1984;41(10):949–58.
2. Institute of Medicine. Improving the quality of healthcare for mental and substance-use conditions: The quality chasm series. Washington, DC: The National Academies Press; 2005.
3. Institute of Medicine. Crossing the quality chasm: A new health system for the 21st century. Washington, DC: The National Academies Press; 2001.
4. **Mokdad AH, Marks JS, Stroup DF, Gerberding JL.** Actual causes of death in the United States, 2000. JAMA. 2004;291(10):1238–45.
5. **Gordon AJ, Sullivan LE, Alford DP, et al.** Update in addiction medicine for the generalist. J Gen Intern Med. 2007;22(8):1190–4.
6. **Gordon AJ, Fiellin DA, Friedmann PD, et al.** Update in addiction medicine for the primary care clinician. J Gen Intern Med. 2008;23 (12):2112–6.
7. **Dunn KM, Saunders KW, Rutter CM, et al.** Opioid prescriptions for chronic pain and overdose: a cohort study. Ann Intern Med. 2010;152 (2):85–92.
8. **Hall AJ, Logan JE, Toblin RL, et al.** Patterns of abuse among unintentional pharmaceutical overdose fatalities. JAMA. 2008;300 (22):2613–20.
9. **Anchersen K, Clausen T, Gossop M, Hansteen V, Waal H.** Prevalence and clinical relevance of corrected QT interval prolongation during methadone and buprenorphine treatment: a mortality assessment study. Addiction. 2009;104(6):993–9.
10. **Paulozzi LJ, Logan JE, Hall AJ, McKinstry E, Kaplan JA, Crosby AE.** A comparison of drug overdose deaths involving methadone and other opioid analgesics in West Virginia. Addiction. 2009;104(9):1541–8.
11. **Krantz MJ, Martin J, Stimmel B, Mehta D, Haigney MC**. QTc Interval Screening in Methadone Treatment: the CSAT Consensus Guideline. Ann Intern Med. 2008.
12. **Gourevitch MN.** First do no harm…Reduction? Ann Intern Med. 2009;150(6):417–8.
13. **McLellan AT, Turner BJ.** Chronic noncancer pain management and opioid overdose: time to change prescribing practices. Ann Intern Med. 2010;152(2):123–4.
14. **Solberg LI, Maciosek MV, Edwards NM.** Primary care intervention to reduce alcohol misuse ranking its health impact and cost effectiveness. Am J Prev Med. 2008;34(2):143–52.
15. **D'Amico EJ, Paddock SM, Burnam A, Kung FY.** Identification of and guidance for problem drinking by general medical providers: results from a national survey. Med Care. 2005;43(3):229–36.
16. **Smith PC, Schmidt SM, Allensworth-Davies D, Saitz R.** Primary care validation of a single-question alcohol screening test. J Gen Intern Med. 2009;24(7):783–8.
17. **Friedmann PD, McCullough D, Chin MH, Saitz R.** Screening and intervention for alcohol problems. A national survey of primary care physicians and psychiatrists. J Gen Intern Med. 2000;15(2):84–91.
18. **Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M.** Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption–II. Addiction. 1993;88(6):791–804.
19. **Allen JP, Litten RZ, Fertig JB, Babor T.** A review of research on the Alcohol Use Disorders Identification Test (AUDIT). Alcoholism: Clinical & Experimental Research. 1997;21(4):613–9.
20. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. Addiction. 2002;97 (9):1183-94.
21. **Vinson DC, Kruse RL, Seale JP.** Simplifying alcohol assessment: two questions to identify alcohol use disorders. Alcohol Clin Exp Res. 2007;31(8):1392–8.
22. **Dawson DA, Grant BF, Stinson FS, Zhou Y.** Effectiveness of the derived Alcohol Use Disorders Identification Test (AUDIT-C) in screening for

alcohol use disorders and risk drinking in the US general population. Alcohol Clin Exp Res. 2005;29(5):844–54.
23. **Gordon AJ, Maisto SA, McNeil M, et al.** Three questions can detect hazardous drinkers. J Fam Pract. 2001;50(4):313–20.
24. **Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA.** alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Arch Intern Med. 1998;158(16):1789–95.
25. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC: American Psychiatric Association; 2000.
26. **Cervantes EA, Miller WR, Tonigan JS.** Comparison of Timeline Follow-Back and Averaging Methods for Quantifying Alcohol Consumption in Treatment Research. Assessment. 1994;1(1):23–30.
27. **Maisto SA, Conigliaro JC, Gordon AJ, Mcginnis KA, Justice AC.** An experimental study of the agreement of self-administration and telephone administration of the Timeline Followback interview. J Stud Alcohol Drugs. 2008;69(3):468–71.
28. **Madras BK, Compton WM, Avula D, Stegbauer T, Stein JB, Clark HW.** Screening, brief interventions, referral to treatment (SBIRT) for illicit drug and alcohol use at multiple healthcare sites: comparison at intake and 6 months later. Drug Alcohol Depend. 2009;99(1–3):280–95.
29. **Krupski A, Sears JM, Joesch JM et al**. Impact of brief interventions and brief treatment on admissions to chemical dependency treatment. Drug Alcohol Depend. 2010.
30. **Babor TF, McRee BG, Kassebaum PA, Grimaldi PL, Ahmed K, Bray J.** Screening, Brief Intervention, and Referral to Treatment (SBIRT): toward a public health approach to the management of substance abuse. Subst Abus. 2007;28(3):7–30.
31. The impact of screening, brief intervention, and referral for treatment on emergency department patients' alcohol use. Ann Emerg Med. 2007;50 (6):699-710, 710.
32. **Fleming MF, Barry KL, Manwell LB, Johnson K, London R.** Brief physician advice for problem alcohol drinkers. A randomized controlled trial in community-based primary care practices. JAMA. 1997;277 (13):1039–45.
33. **Fleming MF, Mundt MP, French MT, Manwell LB, Stauffacher EA, Barry KL.** Brief physician advice for problem drinkers: long-term efficacy and benefit-cost analysis. Alcohol Clin Exp Res. 2002;26(1):36–43.
34. **Otto C, Crackau B, Lohrmann I, et al.** Brief intervention in general hospital for problematic prescription drug use: 12-month outcome. Drug Alcohol Depend. 2009;105(3):221–6.
35. **Zahradnik A, Otto C, Crackau B, et al.** Randomized controlled trial of a brief intervention for problematic prescription drug use in non-treatment-seeking patients. Addiction. 2009;104(1):109–17.
36. **McQueen J, Howe TE, Allan L, Mains D**. Brief interventions for heavy alcohol users admitted to general hospital wards. Cochrane Database Syst Rev. 2009;(3):CD005191.
37. **Guth S, Lindberg SA, Badger GJ, Thomas CS, Rose GL, Helzer JE.** Brief intervention in alcohol-dependent versus nondependent individuals. J Stud Alcohol Drugs. 2008;69(2):243–50.
38. **Stewart SH, Latham PK, Miller PM, Randall P, Anton RF.** Blood pressure reduction during treatment for alcohol dependence: results from the Combining Medications and Behavioral Interventions for Alcoholism (COMBINE) study. Addiction. 2008;103(10):1622–8.
39. **Anton RF, O'Malley SS, Ciraulo DA, et al.** Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE study: a randomized controlled trial. JAMA. 2006;295(17):2003–17.
40. **Nordback I, Pelli H, Lappalainen-Lehto R, Jarvinen S, Raty S, Sand J.** The recurrence of acute alcohol-associated pancreatitis can be reduced: a randomized controlled trial. Gastroenterology. 2009;136(3):848–55.
41. **Gunnell D, Irvine D, Wise L, Davies C, Martin RM.** Varenicline and suicidal behaviour: a cohort study based on data from the General Practice Research Database. BMJ. 2009;339:b3805.
42. **Quinn VP, Hollis JF, Smith KS, et al.** Effectiveness of the 5-As tobacco cessation treatments in nine HMOs. J Gen Intern Med. 2009;24(2):149–54.
43. **Cahill K, Stead LF, Lancaster T**. Nicotine receptor partial agonists for smoking cessation. Cochrane Database Syst Rev. 2008;(3):CD006103.
44. **Kuehn BM.** Studies linking smoking-cessation drug with suicide risk spark concerns. JAMA. 2009;301(10):1007–8.
45. **Kasliwal R, Wilton LV, Shakir SA.** Safety and drug utilization profile of varenicline as used in general practice in England: interim results from a prescription-event monitoring study. Drug Saf. 2009;32(6):499–507.
46. **Walley AY, Alperen JK, Cheng DM, et al.** Office-based management of opioid dependence with buprenorphine: clinical practices and barriers. J Gen Intern Med. 2008;23(9):1393–8.
47. Drug Addiction Treatment Act of 2000, 42 USC§3502a, Drug Addiction Treatment Act of 2000, (2000)
48. **Barry DT, Irwin KS, Jones ES et al**. Integrating Buprenorphine Treatment into Office-based Practice: a Qualitative Study. J Gen Intern Med. 2008.
49. **Soeffing JM, Martin LD, Fingerhood MI, Jasinski DR, Rastegar DA.** Buprenorphine maintenance treatment in a primary care setting: outcomes at 1 year. J Subst Abuse Treat. 2009;37(4):426–30.
50. **Mattick RP, Kimber J, Breen C, Davoli M**. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. Cochrane Database Syst Rev. 2004;(3):CD002207.
51. **Fiellin DA, Moore BA, Sullivan LE, et al.** Long-term treatment with buprenorphine/naloxone in primary care: results at 2-5 years. Am J Addict. 2008;17(2):116–20.
52. **Parran TV, Adelman CA, Merkin B, et al.** Long-term outcomes of office-based buprenorphine/naloxone maintenance therapy. Drug Alcohol Depend. 2010;106(1):56–60.
53. **Fudala PJ, Bridge TP, Herbert S, et al.** Office-based treatment of opiate addiction with a sublingual-tablet formulation of buprenorphine and naloxone. N Engl J Med. 2003;349(10):949–58.
54. **Gordon AJ, Trafton JA, Saxon AJ, et al.** Implementation of buprenorphine in the Veterans Health Administration: Results of the first 3 years. Drug Alcohol Depend. 2007;90:292–6.
55. **Gordon AJ, Liberto J, Granda S, Salmon-Cox S, Andree T, McNicholas L.** Outcomes of DATA 2000 certification trainings for the provision of buprenorphine treatment in the Veterans Health Administration. Am J Addict. 2008;17(6):459–62.
56. **Miller NS, Sheppard LM, Colenda CC, Magen J.** Why physicians are unprepared to treat patients who have alcohol- and drug-related disorders. Acad Med. 2001;76(5):410–8.

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 SEPTEMBER 15, 2011 VOL. 365 NO. 11

Apixaban versus Warfarin in Patients with Atrial Fibrillation

Christopher B. Granger, M.D., John H. Alexander, M.D., M.H.S., John J.V. McMurray, M.D., Renato D. Lopes, M.D., Ph.D., Elaine M. Hylek, M.D., M.P.H., Michael Hanna, M.D., Hussein R. Al-Khalidi, Ph.D., Jack Ansell, M.D., Dan Atar, M.D., Alvaro Avezum, M.D., Ph.D., M. Cecilia Bahit, M.D., Rafael Diaz, M.D., J. Donald Easton, M.D., Justin A. Ezekowitz, M.B., B.Ch., Greg Flaker, M.D., David Garcia, M.D., Margarida Geraldes, Ph.D., Bernard J. Gersh, M.D., Sergey Golitsyn, M.D., Ph.D., Shinya Goto, M.D., Antonio G. Hermosillo, M.D., Stefan H. Hohnloser, M.D., John Horowitz, M.D., Puneet Mohan, M.D., Ph.D., Petr Jansky, M.D., Basil S. Lewis, M.D., Jose Luis Lopez-Sendon, M.D., Prem Pais, M.D., Alexander Parkhomenko, M.D., Freek W.A. Verheugt, M.D., Ph.D., Jun Zhu, M.D., and Lars Wallentin, M.D., Ph.D., for the ARISTOTLE Committees and Investigators*

ABSTRACT

BACKGROUND

Vitamin K antagonists are highly effective in preventing stroke in patients with atrial fibrillation but have several limitations. Apixaban is a novel oral direct factor Xa inhibitor that has been shown to reduce the risk of stroke in a similar population in comparison with aspirin.

METHODS

In this randomized, double-blind trial, we compared apixaban (at a dose of 5 mg twice daily) with warfarin (target international normalized ratio, 2.0 to 3.0) in 18,201 patients with atrial fibrillation and at least one additional risk factor for stroke. The primary outcome was ischemic or hemorrhagic stroke or systemic embolism. The trial was designed to test for noninferiority, with key secondary objectives of testing for superiority with respect to the primary outcome and to the rates of major bleeding and death from any cause.

RESULTS

The median duration of follow-up was 1.8 years. The rate of the primary outcome was 1.27% per year in the apixaban group, as compared with 1.60% per year in the warfarin group (hazard ratio with apixaban, 0.79; 95% confidence interval [CI], 0.66 to 0.95; P<0.001 for noninferiority; P=0.01 for superiority). The rate of major bleeding was 2.13% per year in the apixaban group, as compared with 3.09% per year in the warfarin group (hazard ratio, 0.69; 95% CI, 0.60 to 0.80; P<0.001), and the rates of death from any cause were 3.52% and 3.94%, respectively (hazard ratio, 0.89; 95% CI, 0.80 to 0.99; P=0.047). The rate of hemorrhagic stroke was 0.24% per year in the apixaban group, as compared with 0.47% per year in the warfarin group (hazard ratio, 0.51; 95% CI, 0.35 to 0.75; P<0.001), and the rate of ischemic or uncertain type of stroke was 0.97% per year in the apixaban group and 1.05% per year in the warfarin group (hazard ratio, 0.92; 95% CI, 0.74 to 1.13; P=0.42).

CONCLUSIONS

In patients with atrial fibrillation, apixaban was superior to warfarin in preventing stroke or systemic embolism, caused less bleeding, and resulted in lower mortality. (Funded by Bristol-Myers Squibb and Pfizer; ARISTOTLE ClinicalTrials.gov number, NCT00412984.)

The authors' affiliations are listed in the Appendix. Address reprint requests to Dr. Granger at the Duke Clinical Research Institute, Duke University Medical Center, DUMC Box 3850, Durham, NC 27715, or at christopher.granger@duke.edu.

*The members of the steering committee, as well as other committee members and investigators in the Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) study, are listed in the Supplementary Appendix, available at NEJM.org.

This article (10.1056/NEJMoa1107039) was published on August 28, 2011, and updated on August 30, 2011, at NEJM.org.

N Engl J Med 2011;365:981-92.

Patients with atrial fibrillation are at increased risk for stroke. Warfarin and other vitamin K antagonists are highly effective treatments, reducing the risk of stroke by about two thirds, but their use is limited by a narrow therapeutic range, drug and food interactions, required monitoring, and risk of bleeding.[1] A randomized trial has confirmed the effectiveness of warfarin in the current era.[2] Two new oral anticoagulants have recently been shown to be equivalent or superior to warfarin in preventing stroke or systemic embolism.[3,4] Apixaban is a direct oral factor Xa inhibitor with rapid absorption, a 12-hour half-life, and 25% renal excretion.[5] In patients with atrial fibrillation who were not candidates for vitamin K antagonists, apixaban, as compared with aspirin, reduced the rate of stroke or systemic embolism by 55% without increasing the risk of major bleeding.[6] In the Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) trial,[7] we compared apixaban with warfarin for the prevention of stroke or systemic embolism in patients with atrial fibrillation and at least one additional risk factor for stroke.

Methods

Study Oversight

The trial was designed and led by a steering committee that included academic investigators and representatives of the sponsors (Bristol-Myers Squibb and Pfizer). Approval by the appropriate ethics committees was obtained at all sites. All patients provided written informed consent. The primary analyses were performed both at Bristol-Myers Squibb and at the Duke Clinical Research Institute. All the authors participated in the design of the trial and the planning of the analyses. The first author wrote the first draft of the manuscript, and all the authors participated in subsequent revisions (with no writing assistance other than copy editing) and approved the final version of the manuscript. The trial protocol and statistical analysis plan are available with the full text of this article at NEJM.org. All the authors assume responsibility for the accuracy and completeness of the data and the analyses and for the fidelity of the study to the protocol.

Trial Design

The trial design has been reported previously.[7] With the use of a double-blind, double-dummy design, we randomly assigned patients to treatment with apixaban or dose-adjusted warfarin. The primary objective was to determine whether apixaban was noninferior to warfarin in reducing the rate of stroke (ischemic or hemorrhagic) or systemic embolism among patients with atrial fibrillation and at least one other risk factor for stroke. The primary safety outcome was major bleeding, according to the criteria of the International Society on Thrombosis and Haemostasis (ISTH). Key secondary objectives were to determine whether apixaban was superior to warfarin with respect to the primary outcome and to the rates of major bleeding and death from any cause. To control the overall type I error, prespecified hierarchical sequential testing was performed first on the primary outcome for noninferiority, then on the primary outcome for superiority, then on major bleeding, and finally on death from any cause.

Study Population

Eligible patients had atrial fibrillation or flutter at enrollment or two or more episodes of atrial fibrillation or flutter, as documented by electrocardiography, at least 2 weeks apart in the 12 months before enrollment. In addition, at least one of the following risk factors for stroke was required: an age of at least 75 years; previous stroke, transient ischemic attack, or systemic embolism; symptomatic heart failure within the previous 3 months or left ventricular ejection fraction of no more than 40%; diabetes mellitus; or hypertension requiring pharmacologic treatment. Key exclusion criteria were atrial fibrillation due to a reversible cause, moderate or severe mitral stenosis, conditions other than atrial fibrillation that required anticoagulation (e.g., a prosthetic heart valve), stroke within the previous 7 days, a need for aspirin at a dose of >165 mg a day or for both aspirin and clopidogrel, and severe renal insufficiency (serum creatinine level of >2.5 mg per deciliter [221 μmol per liter] or calculated creatinine clearance of <25 ml per minute).[7] Patients were classified as not having received warfarin previously if they had used warfarin or another vitamin K antagonist for no more than 30 consecutive days. Investigators at

all study centers were encouraged to enroll a sizable proportion of patients (≥40%) who had not previously received warfarin.

RANDOMIZATION AND STUDY DRUGS

Randomization was stratified according to whether patients had received warfarin previously and according to clinical site. Apixaban or matching placebo was administered twice daily, with apixaban given in 5-mg doses; 2.5-mg doses were used in a subset of patients with two or more of the following criteria: an age of at least 80 years, a body weight of no more than 60 kg, or a serum creatinine level of 1.5 mg per deciliter (133 μmol per liter) or more. Warfarin (or matching placebo) was provided as 2-mg tablets and was adjusted to achieve a target international normalized ratio (INR) of 2.0 to 3.0. Patients who were receiving a vitamin K antagonist before randomization were instructed to discontinue the drug 3 days before randomization, and the study drug was initiated when the INR was less than 2.0. INRs were monitored with the use of a blinded, encrypted, point-of-care INR device. An algorithm was provided to guide the adjustment of the warfarin dose. The time that patients' INRs were within the therapeutic range was calculated by the Rosendaal method.[8] A program was implemented to improve the quality of INR control through education and feedback at the site and country levels.

An algorithm was provided to manage temporary discontinuations of the study drug around the time of interventional procedures while maintaining concealment of the group assignments. At the end of the trial, when patients discontinued the study drug, guidance was provided in making the transition to open-label warfarin while maintaining concealment of the treatment assignments and ensuring appropriate anticoagulation. In addition to monthly study visits focusing on control of the INR, visits every 3 months included an assessment of clinical outcomes and adverse events. For each patient who was lost to follow-up or who withdrew consent, attempts were made to determine vital status at the end of the trial.

STUDY OUTCOMES

The primary efficacy outcome was stroke or systemic embolism. Stroke was defined as a focal neurologic deficit, from a nontraumatic cause, lasting at least 24 hours and was categorized as ischemic (with or without hemorrhagic transformation), hemorrhagic, or of uncertain type (in the case of patients who did not undergo brain imaging or in whom an autopsy was not performed). The key secondary efficacy outcome was death from any cause. The rate of myocardial infarction was also assessed as a secondary efficacy outcome.

The primary safety outcome was major bleeding, which was defined, according to the ISTH criteria,[9] as clinically overt bleeding accompanied by a decrease in the hemoglobin level of at least 2 g per deciliter or transfusion of at least 2 units of packed red cells, occurring at a critical site, or resulting in death. The secondary safety outcome was a composite of major bleeding and clinically relevant nonmajor bleeding, which was defined as clinically overt bleeding that did not satisfy the criteria for major bleeding and that led to hospital admission, physician-guided medical or surgical treatment, or a change in antithrombotic therapy. Other safety outcomes included any bleeding, other adverse events, and liver-function abnormalities.

The primary and secondary efficacy and safety outcomes were adjudicated on the basis of prespecified criteria by a clinical-events committee whose members were not aware of study-group assignments. For details, see the Supplementary Appendix, available at NEJM.org.

STATISTICAL ANALYSIS

The primary noninferiority hypothesis required that apixaban preserve at least 50% of the relative reduction in the risk of stroke or systemic embolism associated with warfarin (62%) in six previous, major randomized, controlled trials.[10] This hypothesis provided a lower 95% confidence interval of 1.88 for the relative risk with placebo as compared with warfarin, and one half of this value was 1.44 (or 1.38 on a log scale). We estimated that with the occurrence of the primary outcome in 448 patients, the study would have 90% power to ensure that the upper boundary of the 99% confidence interval for the relative risk would be less than 1.44 and that the upper boundary of the 95% confidence interval for the relative risk would be less than 1.38, on the assumption that apixaban and warfarin had identical effects. On the basis of the overall event rate during the trial, we planned to recruit 18,000 pa-

tients in order to reach this number of events with approximately 2 years of follow-up. An independent data and safety monitoring committee reviewed the accumulating trial data, with one prespecified interim analysis for efficacy.

The primary and key secondary analyses were performed with the use of the Cox proportional-hazards model, with previous warfarin status and geographic region (North America, South America, Europe, or Asian Pacific) used as strata in the model. The primary and secondary efficacy analyses included all patients who underwent randomization (intention-to-treat population) and included all events from the time of randomization until the cutoff date for efficacy outcomes (predefined as January 30, 2011). The analyses of bleeding events included all patients who received at least one dose of a study drug and included all events from the time the first dose of a study drug was received until 2 days after the last dose was received. In a modified intention-to-treat sensitivity analysis, we analyzed bleeding events that occurred in patients who received at least one dose of a study drug and included all

Table 1. Baseline Characteristics of the Patients.*

Characteristic	Apixaban (N=9120)	Warfarin (N=9081)
Age — yr		
Median	70	70
Interquartile range	63–76	63–76
Female sex — no. (%)	3234 (35.5)	3182 (35.0)
Region — no. (%)		
North America	2249 (24.7)	2225 (24.5)
Latin America	1743 (19.1)	1725 (19.0)
Europe	3672 (40.3)	3671 (40.4)
Asian Pacific	1456 (16.0)	1460 (16.1)
Systolic blood pressure — mm Hg		
Median	130	130
Interquartile range	120–140	120–140
Weight — kg		
Median	82	82
Interquartile range	70–96	70–95
Prior myocardial infarction — no. (%)	1319 (14.5)	1266 (13.9)
Prior clinically relevant or spontaneous bleeding — no. (%)	1525 (16.7)	1515 (16.7)
History of fall within previous year — no. (%)	386 (4.2)	367 (4.0)
Type of atrial fibrillation — no. (%)		
Paroxysmal	1374 (15.1)	1412 (15.5)
Persistent or permanent	7744 (84.9)	7668 (84.4)
Prior use of vitamin K antagonist for >30 consecutive days — no. (%)	5208 (57.1)	5193 (57.2)
Qualifying risk factors		
Age ≥75 yr — no. (%)	2850 (31.2)	2828 (31.1)
Prior stroke, TIA, or systemic embolism — no. (%)	1748 (19.2)	1790 (19.7)
Heart failure or reduced left ventricular ejection fraction — no. (%)	3235 (35.5)	3216 (35.4)
Diabetes — no. (%)	2284 (25.0)	2263 (24.9)
Hypertension requiring treatment — no. (%)	7962 (87.3)	7954 (87.6)
$CHADS_2$ score		
Mean	2.1±1.1	2.1±1.1
Distribution — no. (%)		
1	3100 (34.0)	3083 (34.0)
2	3262 (35.8)	3254 (35.8)
≥3	2758 (30.2)	2744 (30.2)

Table 1. (Continued.)

Characteristic	Apixaban (N=9120)	Warfarin (N=9081)
Medications at time of randomization — no. (%)		
ACE inhibitor or ARB	6464 (70.9)	6368 (70.1)
Amiodarone	1009 (11.1)	1042 (11.5)
Beta-blocker	5797 (63.6)	5685 (62.6)
Aspirin	2859 (31.3)	2773 (30.5)
Clopidogrel	170 (1.9)	168 (1.9)
Digoxin	2916 (32.0)	2912 (32.1)
Calcium blocker	2744 (30.1)	2823 (31.1)
Statin	4104 (45.0)	4095 (45.1)
Nonsteroidal antiinflammatory agent	752 (8.2)	768 (8.5)
Gastric antacid drugs	1683 (18.5)	1667 (18.4)
Renal function, creatinine clearance — no. (%)		
Normal, >80 ml/min	3761 (41.2)	3757 (41.4)
Mild impairment, >50 to 80 ml/min	3817 (41.9)	3770 (41.5)
Moderate impairment (>30 to 50 ml/min)	1365 (15.0)	1382 (15.2)
Severe impairment (≤30 ml/min)	137 (1.5)	133 (1.5)
Not reported	40 (0.4)	39 (0.4)

* Plus–minus values are means ±SD. None of the characteristics differed significantly between the groups (P>0.05 for all comparisons). ACE denotes angiotensin-converting enzyme, ARB angiotensin-receptor blocker, and TIA transient ischemic attack.

Table 2. Efficacy Outcomes.*

Outcome	Apixaban Group (N=9120)		Warfarin Group (N=9081)		Hazard Ratio (95% CI)	P Value
	Patients with Event	Event Rate	Patients with Event	Event Rate		
	no.	*%/yr*	*no.*	*%/yr*		
Primary outcome: stroke or systemic embolism	212	1.27	265	1.60	0.79 (0.66–0.95)	0.01
Stroke	199	1.19	250	1.51	0.79 (0.65–0.95)	0.01
Ischemic or uncertain type of stroke	162	0.97	175	1.05	0.92 (0.74–1.13)	0.42
Hemorrhagic stroke	40	0.24	78	0.47	0.51 (0.35–0.75)	<0.001
Systemic embolism	15	0.09	17	0.10	0.87 (0.44–1.75)	0.70
Key secondary efficacy outcome: death from any cause	603	3.52	669	3.94	0.89 (0.80–0.998)	0.047
Other secondary outcomes						
Stroke, systemic embolism, or death from any cause	752	4.49	837	5.04	0.89 (0.81–0.98)	0.02
Myocardial infarction	90	0.53	102	0.61	0.88 (0.66–1.17)	0.37
Stroke, systemic embolism, myocardial infarction, or death from any cause	810	4.85	906	5.49	0.88 (0.80–0.97)	0.01
Pulmonary embolism or deep-vein thrombosis	7	0.04	9	0.05	0.78 (0.29–2.10)	0.63

* Analyses were performed on data from the intention-to-treat population and included all events through the cutoff date for efficacy outcomes of January 30, 2011; comparisons of the primary outcome and of death from any cause were analyzed as part of hierarchical sequence testing (starting with testing the primary outcome for noninferiority, then the primary outcome for superiority, then major bleeding, and finally death from any cause), to control the type I error.

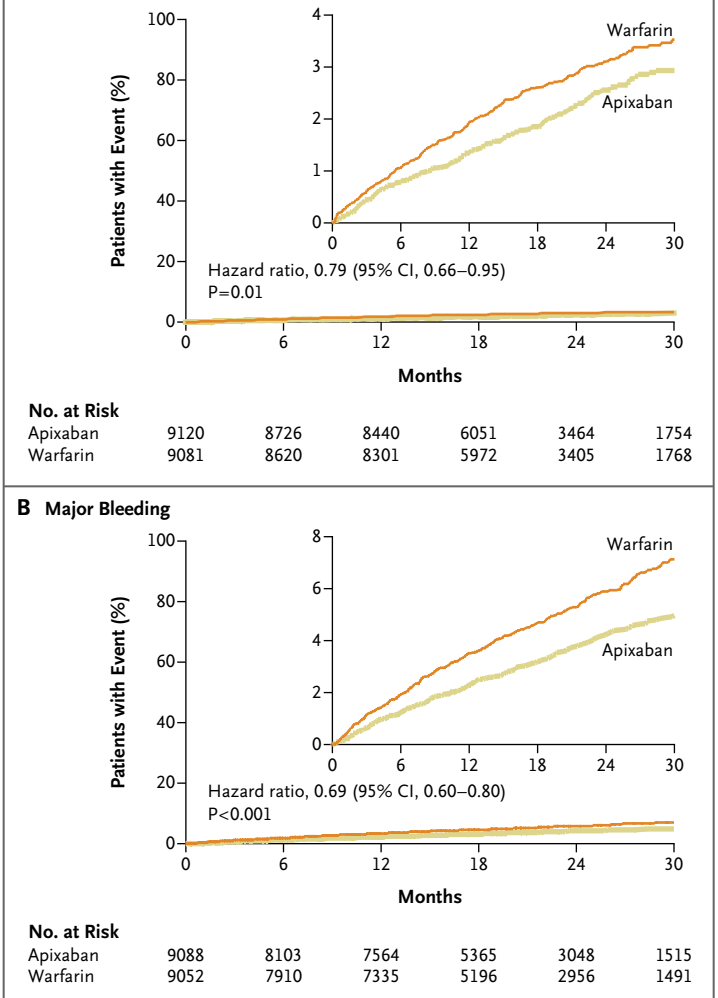


Figure 1. Kaplan–Meier Curves for the Primary Efficacy and Safety Outcomes.
The primary efficacy outcome (Panel A) was stroke or systemic embolism. The primary safety outcome (Panel B) was major bleeding, as defined according to the criteria of the International Society on Thrombosis and Haemostasis. The inset in each panel shows the same data on an enlarged segment of the y axis.

events from the time of randomization until January 30, 2011. All reported P values for noninferiority are one-sided, and all reported P values for superiority are two-sided. All statistical analyses were performed with the use of SAS software, version 9.0 (SAS Institute).

RESULTS

PATIENTS AND FOLLOW-UP

From December 19, 2006, through April 2, 2010, we recruited 18,201 patients at 1034 clinical sites in 39 countries. A total of 9120 were assigned to the apixaban group and 9081 to the warfarin group. The two groups were well balanced with respect to baseline characteristics (Table 1). The median age was 70 years; 35.3% of the patients were women, and the mean $CHADS_2$ score was 2.1. (The $CHADS_2$ score, an index of the risk of stroke in patients with atrial fibrillation, ranges from 1 to 6, with higher scores indicating a greater risk of stroke.) Approximately 57% of the patients had previously received a vitamin K antagonist, and 19% had had a previous stroke, transient ischemic attack, or systemic embolism.

Data on vital status at the end of the trial were missing for 380 patients (2.1%). The absence of data on vital status was due to withdrawal of consent in the case of 92 patients in the apixaban group (1.0%) and 107 patients in the warfarin group (1.2%) and was due to loss to follow-up in the case of 35 patients in the apixaban group (0.4%) and 34 in the warfarin group (0.4%).

STUDY DRUGS

A reduced dose of apixaban (2.5 mg twice daily) or placebo was administered in 428 patients in the apixaban group (4.7%) and 403 in the warfarin group (4.4%). Fewer patients in the apixaban group than in the warfarin group discontinued a study drug before the end of the study: 25.3% of the patients in the apixaban group, with 3.6% of the discontinuations due to death, versus 27.5% of patients in the warfarin group, with 3.8% due to death (P=0.001). Patients in the warfarin group had an INR in the therapeutic range (2.0 to 3.0) for a median of 66.0% of the time and a mean of 62.2% of the time, after the exclusion of INR values during the first 7 days after randomization and during study-drug interruptions.

PRIMARY OUTCOME

The primary outcome of stroke or systemic embolism occurred in 212 patients in the apixaban group (1.27% per year) as compared with 265 patients in the warfarin group (1.60% per year) (hazard ratio in the apixaban group, 0.79; 95% confidence interval [CI], 0.66 to 0.95; P<0.001 for noninferiority and P=0.01 for superiority) (Table 2 and Fig. 1A). The rate of hemorrhagic stroke was 49% lower in the apixaban group than in the warfarin group, and the rate of ischemic or uncertain type of stroke was 8% lower in the apixaban group than in the warfarin group (Table 2). Fatal or disabling stroke occurred in 84 patients in the apixaban group (0.50% per year) as com-

Table 3. Bleeding Outcomes and Net Clinical Outcomes.*

Outcome	Apixaban Group (N=9088)		Warfarin Group (N=9052)		Hazard Ratio (95% CI)	P Value
	Patients with Event	Event Rate	Patients with Event	Event Rate		
	no.	*%/yr*	*no.*	*%/yr*		
Primary safety outcome: ISTH major bleeding†	327	2.13	462	3.09	0.69 (0.60–0.80)	<0.001
Intracranial	52	0.33	122	0.80	0.42 (0.30–0.58)	<0.001
Other location	275	1.79	340	2.27	0.79 (0.68–0.93)	0.004
Gastrointestinal	105	0.76	119	0.86	0.89 (0.70–1.15)	0.37
Major or clinically relevant nonmajor bleeding	613	4.07	877	6.01	0.68 (0.61–0.75)	<0.001
GUSTO severe bleeding	80	0.52	172	1.13	0.46 (0.35–0.60)	<0.001
GUSTO moderate or severe bleeding	199	1.29	328	2.18	0.60 (0.50–0.71)	<0.001
TIMI major bleeding	148	0.96	256	1.69	0.57 (0.46–0.70)	<0.001
TIMI major or minor bleeding	239	1.55	370	2.46	0.63 (0.54–0.75)	<0.001
Any bleeding	2356	18.1	3060	25.8	0.71 (0.68–0.75)	<0.001
Net clinical outcomes						
Stroke, systemic embolism, or major bleeding	521	3.17	666	4.11	0.77 (0.69–0.86)	<0.001
Stroke, systemic embolism, major bleeding, or death from any cause	1009	6.13	1168	7.20	0.85 (0.78–0.92)	<0.001

* The bleeding outcomes were assessed in patients who received at least one dose of a study drug and events that occurred from the time the patients received the first dose of the study drug through 2 days after they received the last dose. The net clinical outcome includes all efficacy outcomes through the cutoff date for the efficacy analysis and bleeding outcomes that occurred from the time the patients received the first dose of the study drug through 2 days after they received the last dose. GUSTO denotes Global Use of Strategies to Open Occluded Coronary Arteries, and TIMI Thrombolysis in Myocardial Infarction.

† The comparison of the primary safety outcome of bleeding according to the International Society on Thrombosis and Haemostasis (ISTH) criteria is in the hierarchical sequence preserving a type I error.

pared with 117 patients in the warfarin group (0.71% per year) (hazard ratio, 0.71; 95% CI, 0.54 to 0.94). Ischemic stroke occurred in 149 patients in the apixaban group and in 155 patients in the warfarin group, and an unknown type of stroke occurred in 14 patients in the apixaban group and 21 patients in the warfarin group. Among the patients with ischemic strokes, hemorrhagic transformation occurred in 12 patients in the apixaban group and 20 patients in the warfarin group. Fatal stroke occurred in 42 patients in the apixaban group and 67 patients in the warfarin group.

KEY SECONDARY AND OTHER EFFICACY OUTCOMES

The rate of death from any cause was lower in the apixaban group than in the warfarin group (3.52% per year vs. 3.94% per year; hazard ratio, 0.89; 95% CI, 0.80 to 0.99; P=0.047). The rate of death from cardiovascular causes (including death from hemorrhagic stroke) was 1.80% per year in the apixaban group and 2.02% per year in the warfarin group (hazard ratio, 0.89; 95% CI, 0.76 to 1.04), and the rate of death from noncardiovascular causes (including fatal bleeding other than that from hemorrhagic stroke) was 1.14% per year in the apixaban group and 1.22% per year in the warfarin group (hazard ratio, 0.93; 95% CI, 0.77 to 1.13). The rate of myocardial infarction was lower in the apixaban group than in the warfarin group, but the difference was not significant (Table 2).

BLEEDING

Major bleeding, as defined according to ISTH criteria, occurred in 327 patients in the apixaban group (2.13% per year), as compared with 462 patients in the warfarin group (3.09% per year) (hazard ratio, 0.69; 95% CI, 0.60 to 0.80; P<0.001) (Table 3 and Fig. 1B). There appeared to be an even greater reduction in the rate of serious bleeding as defined according to the Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO) criteria for severe bleeding and according to the Thrombolysis in Myocardial Infarction (TIMI) criteria for major bleeding (Table 3). The

A Primary Efficacy Outcome: Stroke and Systemic Embolism

Subgroup	No. of Patients	Apixaban	Warfarin	P Value for Interaction
		no. of events (%/yr)		
All patients	18,201	212 (1.27)	265 (1.60)	
Prior use of warfarin or other vitamin K antagonist				0.39
Yes	10,401	102 (1.1)	138 (1.5)	
No	7,800	110 (1.5)	127 (1.8)	
Age				0.12
<65 yr	5,471	51 (1.0)	44 (0.9)	
65 to <75 yr	7,052	82 (1.3)	112 (1.7)	
≥75 yr	5,678	79 (1.6)	109 (2.2)	
Sex				0.60
Male	11,785	132 (1.2)	160 (1.5)	
Female	6,416	80 (1.4)	105 (1.8)	
Weight				0.26
≤60 kg	1,985	34 (2.0)	52 (3.2)	
>60 kg	16,154	177 (1.2)	212 (1.4)	
Type of atrial fibrillation				0.70
Permanent or persistent	15,412	191 (1.4)	235 (1.7)	
Paroxysmal	2,786	21 (0.8)	30 (1.1)	
Prior stroke or TIA				0.71
Yes	3,436	73 (2.5)	98 (3.2)	
No	14,765	139 (1.0)	167 (1.2)	
Diabetes mellitus				0.71
Yes	4,547	57 (1.4)	75 (1.9)	
No	13,654	155 (1.2)	190 (1.5)	
Heart failure				0.50
Yes	5,541	70 (1.4)	79 (1.6)	
No	12,660	142 (1.2)	186 (1.6)	
$CHADS_2$ score				0.45
1	6,183	44 (0.7)	51 (0.9)	
2	6,516	74 (1.2)	82 (1.4)	
≥3	5,502	94 (1.9)	132 (2.8)	
Level of renal impairment				0.72
Severe or moderate	3,017	54 (2.1)	69 (2.7)	
Mild	7,587	87 (1.2)	116 (1.7)	
No impairment	7,518	70 (1.0)	79 (1.1)	
Apixaban dose				0.22
2.5 mg twice daily or placebo	831	12 (1.7)	22 (3.3)	
5 mg twice daily or placebo	17,370	200 (1.3)	243 (1.5)	
Geographic region				0.44
North America	4,474	42 (1.0)	56 (1.3)	
Latin America	3,468	43 (1.4)	52 (1.8)	
Europe	7,343	75 (1.1)	77 (1.1)	
Asian Pacific	2,916	52 (2.0)	80 (3.1)	
Aspirin use at randomization				0.44
Yes	5,632	70 (1.3)	94 (1.9)	
No	12,569	142 (1.2)	171 (1.5)	

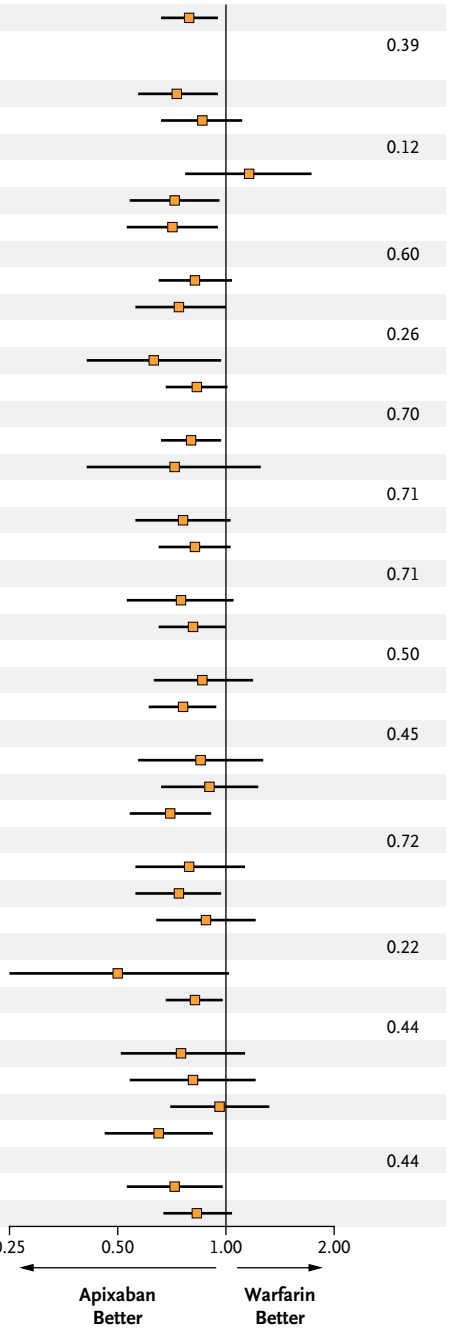
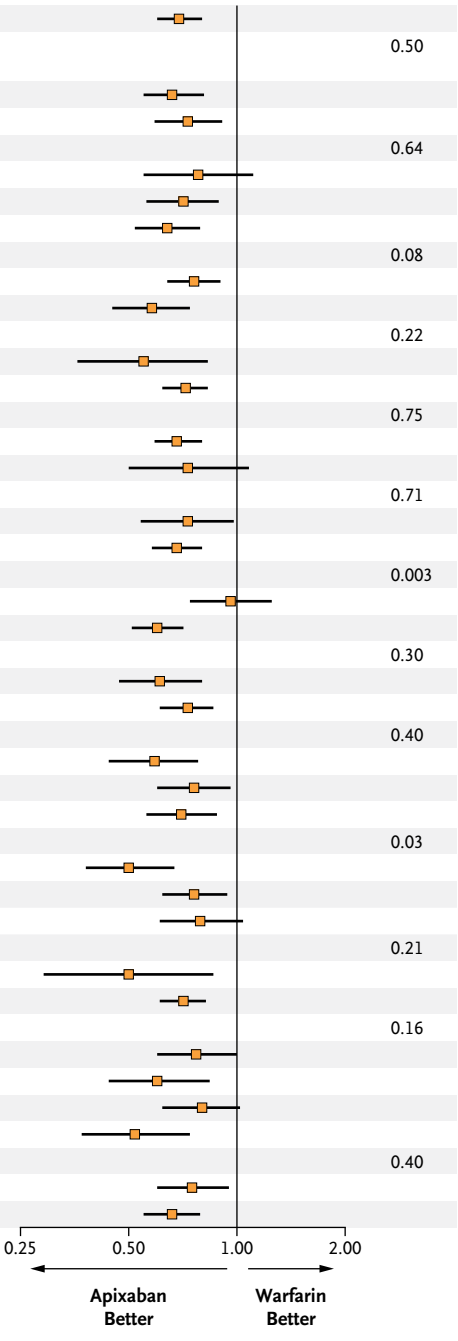


Figure 2. Relative Risks of the Primary Efficacy and Safety Outcomes, According to Major Prespecified Subgroups.

Prespecified subgroups not included in the figure were subgroups according to race, ethnic group, body-mass index, number of risk factors, age of 75 years or more, and use or nonuse of clopidogrel at the time of randomization, as well as subgroups of women according to age group. Heart failure was defined as symptomatic heart failure or a left ventricular ejection fraction of 40% or less. The $CHADS_2$ score, an index of the risk of stroke in patients with atrial fibrillation, ranges from 1 to 6, with higher scores indicating a greater risk of stroke. TIA denotes transient ischemic attack.

B Major Bleeding

Subgroup	No. of Patients	Apixaban	Warfarin	P Value for Interaction
		no. of events (%/yr)		
All patients	18,140	327 (2.13)	462 (3.09)	
Prior use of warfarin or other vitamin K antagonist				0.50
Yes	10,376	185 (2.1)	274 (3.2)	
No	7,764	142 (2.2)	188 (3.0)	
Age				0.64
<65 yr	5,455	56 (1.2)	72 (1.5)	
65 to <75 yr	7,030	120 (2.0)	166 (2.8)	
≥75 yr	5,655	151 (3.3)	224 (5.2)	
Sex				0.08
Male	11,747	225 (2.3)	294 (3.0)	
Female	6,393	102 (1.9)	168 (3.3)	
Weight				0.22
≤60 kg	1,978	36 (2.3)	62 (4.3)	
>60 kg	16,102	290 (2.1)	398 (3.0)	
Type of atrial fibrillation				0.75
Permanent or persistent	15,361	283 (2.2)	402 (3.2)	
Paroxysmal	2,776	44 (1.9)	60 (2.6)	
Prior stroke or TIA				0.71
Yes	3,422	77 (2.8)	106 (3.9)	
No	14,718	250 (2.0)	356 (2.9)	
Diabetes mellitus				0.003
Yes	4,526	112 (3.0)	114 (3.1)	
No	13,614	215 (1.9)	348 (3.1)	
Heart failure				0.30
Yes	5,527	87 (1.9)	137 (3.1)	
No	12,613	240 (2.2)	325 (3.1)	
$CHADS_2$ score				0.40
1	6,169	76 (1.4)	126 (2.3)	
2	6,492	125 (2.3)	163 (3.0)	
≥3	5,479	126 (2.9)	173 (4.2)	
Level of renal impairment				0.03
Severe or moderate	3,005	73 (3.2)	142 (6.4)	
Mild	7,565	157 (2.5)	199 (3.2)	
No impairment	7,496	96 (1.5)	119 (1.8)	
Apixaban dose				0.21
2.5 mg twice daily or placebo	826	20 (3.3)	37 (6.7)	
5 mg twice daily or placebo	17,314	307 (2.1)	425 (3.0)	
Geographic region				0.16
North America	4,463	106 (2.8)	137 (3.6)	
Latin America	3,460	60 (2.1)	94 (3.5)	
Europe	7,313	110 (1.7)	135 (2.2)	
Asian Pacific	2,904	51 (2.1)	96 (4.1)	
Aspirin use at randomization				0.40
Yes	5,608	129 (2.7)	164 (3.7)	
No	12,532	198 (1.9)	298 (2.8)	



rate of intracranial hemorrhage was 0.33% per year in the apixaban group and 0.80% per year in the warfarin group (hazard ratio, 0.42; 95% CI, 0.30 to 0.58; P<0.001), and the rate of any bleeding was 25.8% per year in the warfarin group and 18.1% per year in the apixaban group, an absolute reduction of 7.7 percentage points (P<0.001). In a modified intention-to-treat sensitivity analysis that included the entire treatment period, there was a consistent 27% relative reduction in

the rate of major bleeding in the apixaban group, as compared with the warfarin group (P<0.001). Fatal bleeding (including fatal hemorrhagic stroke), as evaluated in the intention-to-treat analysis, occurred in 34 patients in the apixaban group and 55 patients in the warfarin group.

SUBGROUPS

The reduction in the primary outcome with apixaban was consistent across all major subgroups (Fig. 2), and statistical tests for interaction were not significant (P>0.10) for all of the 21 predefined subgroups. With respect to the outcome of major bleeding, the only baseline characteristics for which the interaction was significant were diabetes status and renal function, with a greater reduction in bleeding among patients who did not have diabetes (P=0.003 for interaction) and among patients with moderate or severe renal impairment (P=0.03 for interaction).

OVERALL SAFETY OUTCOMES

Adverse events occurred in almost equal proportions of patients in the apixaban group and in the warfarin group (81.5% of the patients in the apixaban group and 83.1% of patients in the warfarin group), as did serious adverse events (35.0% and 36.5% in the two groups, respectively) (for details, see the Supplementary Appendix). The rates of abnormalities on liver-function testing and liver-related serious adverse events were similar in the two groups.

DISCUSSION

In patients with atrial fibrillation and at least one additional risk factor for stroke, the use of apixaban, as compared with warfarin, significantly reduced the risk of stroke or systemic embolism by 21%, major bleeding by 31%, and death by 11%. For every 1000 patients treated for 1.8 years, apixaban, as compared with warfarin, prevented a stroke in 6 patients, major bleeding in 15 patients, and death in 8 patients. The predominant effect on stroke prevention was on hemorrhagic stroke, with prevention of a hemorrhagic stroke in 4 patients per 1000 and prevention of an ischemic or unknown type of stroke in 2 patients per 1000.

The results were consistent in subgroups according to geographic region, status with respect to previous warfarin exposure, age, sex, level of renal impairment, and risk factors for stroke, as well as in other predefined subgroups. Apixaban had an acceptable side-effect profile, with no unexpected side effects, and the rate of discontinuation of the study drug was lower in the apixaban group than in the warfarin group.

Warfarin is highly effective in preventing stroke in patients with atrial fibrillation but is associated with a variable response, has drug and food interactions, requires regular monitoring for dose adjustment, and carries a risk of bleeding (including intracranial hemorrhage). In part because of these limitations, only about half of patients who would benefit from warfarin therapy actually receive the drug.[11] The alternative treatment regimen with apixaban (at a dose of 5 mg twice daily), which does not require anticoagulation monitoring, not only is more effective than warfarin for stroke prevention but also accomplishes this goal at a substantially lower risk of bleeding and with lower rates of discontinuation. These findings are supported by the results of the Apixaban Versus Acetylsalicylic Acid [ASA] to Prevent Stroke in Atrial Fibrillation Patients Who Have Failed or Are Unsuitable for Vitamin K Antagonist Treatment trial (AVERROES; ClinicalTrials.gov number, NCT00496769),[6] in which the same apixaban regimen, as compared with low-dose aspirin, was shown to substantially reduce the risk of stroke without any difference in the rates of major bleeding and with lower rates of discontinuation. Although major bleeding was less common with apixaban, at a dose of 5 mg twice daily, than with warfarin in patients with atrial fibrillation, the use of the same dose of apixaban, as compared with placebo, resulted in more bleeding in patients with acute coronary syndromes who were receiving both aspirin and clopidogrel.[12] The significant reduction in mortality observed in our study was consistent with trends toward lower rates of death among patients receiving apixaban than among those receiving aspirin in the AVERROES trial.

Two alternative oral anticoagulants, the direct thrombin inhibitor dabigatran[3] and the factor Xa inhibitor rivaroxaban,[4] have recently been shown in randomized clinical trials to be at least as effective as warfarin in preventing stroke. Each of these agents, like apixaban, has the major advantage of convenience, since there is no need for anticoagulation monitoring. In the Randomized Evaluation of Long-Term Anticoagulation Therapy trial (RE-LY, NCT00262600) the oral di-

rect thrombin inhibitor dabigatran administered in two doses per day was compared with open-label warfarin. The 150-mg dose of dabigatran administered twice daily, as compared with warfarin, was shown to reduce the rate of stroke, including ischemic or unspecified stroke, with a similar overall rate of bleeding, although the rate of gastrointestinal bleeding was increased. The 110-mg dose administered twice daily was associated with a rate of stroke that was similar to that with warfarin but with a lower rate of major bleeding. Both doses resulted in lower rates of intracranial hemorrhage. In our study, apixaban at a dose of 5 mg twice daily (with the recommendation to use a reduced dose in patients with a predicted higher drug exposure) appears to combine the advantages of each of the two doses of dabigatran, with both a greater overall reduction in the rate of stroke and a lower rate of bleeding than the rates with warfarin. As compared with warfarin, apixaban is also associated with a reduction in the rate of gastrointestinal bleeding and with consistently lower bleeding rates across age groups[13] and all other major subgroups. Fewer patients receiving apixaban had a myocardial infarction than those receiving either warfarin (in our study) or aspirin (in the AVERROES trial).

Rivaroxaban, the second new alternative, was shown to be noninferior to warfarin for the prevention of stroke and systemic embolism in the intention-to-treat analysis in the Rivaroxaban Once Daily Oral Direct Factor Xa Inhibition Compared with Vitamin K Antagonism for Prevention of Stroke and Embolism Trial in Atrial Fibrillation (ROCKET AF, NCT00403767).[4] The rates of intracranial hemorrhage and fatal bleeding were lower with rivaroxaban than with warfarin, but there was no advantage with respect to other major bleeding. The differences between our findings and those of other trials comparing novel anticoagulants with warfarin may be related to differences in the doses of drugs, the pharmacokinetic and pharmacodynamic properties of the drugs,[14] patient populations, or other features of the clinical-trial design. The lower risk of hemorrhagic stroke associated with all three novel anticoagulants suggests that there is a specific risk associated with warfarin, possibly related to its inhibition of multiple coagulation factors or interaction between warfarin and tissue factor VIIa complexes in the brain.[15]

In conclusion, in patients with atrial fibrillation, apixaban was superior to warfarin in preventing stroke or systemic embolism, caused less bleeding, and resulted in lower mortality.

Supported by Bristol-Myers Squibb and Pfizer.

Dr. Granger reports receiving grants from AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, the Medtronic Foundation, Merck, Sanofi-Aventis, Astellas, and the Medicines Company, consulting fees from AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Hoffmann-La Roche, Novartis, Otsuka Pharmaceutical, Sanofi-Aventis, and the Medicines Company, and support from the Medtronic Foundation and Merck for travel, accommodations, or meeting expenses; Dr. Alexander, receiving grants from Merck/Schering-Plough and Regado Biosciences and consulting fees from Merck/Schering-Plough, AstraZeneca, Boehringer Ingelheim, Ortho-McNeil-Janssen, PolyMedix, Regado Biosciences, and Bayer; Dr. Lopes, receiving grants from Bristol-Myers Squibb, AstraZeneca, Boehringer Ingelheim, and Daiichi Sankyo and consulting fees from Bristol-Myers Squibb; Dr. Hylek, receiving consulting fees from Johnson & Johnson, Boehringer Ingelheim, Bristol-Myers Squibb, Daiichi Sankyo, Merck, Pfizer, and Ortho-McNeil-Janssen and lecture fees from Boehringer Ingelheim; Dr. Hanna, being an employee of Bristol-Myers Squibb and receiving stock as part of his compensation; Dr. Ansell, being a member of Bristol-Myers Squibb's data and safety monitoring boards, receiving consulting fees from Ortho-McNeil-Janssen, Daiichi Sankyo, Boehringer Ingelheim, and Bristol-Myers Squibb, and receiving payments from Daiichi Sankyo for developing educational presentations; Dr. Atar, receiving consulting fees or honoraria from Bristol-Myers Squibb; Dr. Avezum, receiving consulting fees from Boehringer Ingelheim and GlaxoSmithKline; Dr. Bahit, receiving consulting fees or honoraria and other research support from Bristol-Myers Squibb; Dr. Diaz, receiving consulting fees or honoraria and other research support from Bristol-Myers Squibb; Dr. Ezekowitz, receiving research support, honoraria, and payments for continuing medical education events from AstraZeneca, Amgen, Abbott, Servier, Johnson & Johnson, Pfizer, and Bristol-Myers Squibb; Dr. Flaker, receiving consulting fees from Boehringer Ingelheim and Sanofi-Aventis and grants from Boehringer Ingelheim and Sanofi-Aventis; Dr. Garcia, receiving consulting fees from Bristol-Myers Squibb, Pfizer, Boehringer Ingelheim, and Daiichi Sankyo and payments from Boehringer Ingelheim for developing educational presentations; Dr. Geraldes, being an employee of Bristol-Myers Squibb and receiving stock or stock options annually; Dr. Gersh, receiving consulting fees from Ortho-McNeil-Janssen, Amorcyte, Abbott Laboratories, GE Healthcare, St. Jude Medical, Medispec, Merck, and Boston Scientific; Dr. Goto, being a board member of Bristol-Myers Squibb and Sanofi-Aventis, receiving grants from Boehringer Ingelheim, Otuska, Eisai, Sanofi-Aventis, and Daiichi Sankyo, consulting fees from Eisai, lecture fees from Eisai, Otsuka, Daiichi Sankyo, Sanofi-Aventis, Bayer, Novartis, AstraZeneca, Asteras, Pfizer, Medtronics-Japan, Tanabe-Mitsubishi, Takeda, Mochida, and MSD, and payments from Bayer and Sanofi-Aventis for developing educational presentations; Dr. Hermosillo, receiving consulting fees or honoraria from Bristol-Myers Squibb; Dr. Hohnloser, receiving consulting fees from Sanofi-Aventis, St. Jude Medical, Boehringer Ingelheim, Cardiome, and Medtronic Vascular and lecture fees from Sanofi-Aventis, Bristol-Myers Squibb, Pfizer, Boehringer Ingelheim, and St. Jude Medical; Dr. Horowitz, receiving research support from Bristol-Myers Squibb; Dr. Mohan, being an employee of Bristol-Myers Squibb and receiving performance-based stock or stock options annually; Dr. Lewis, being an advisory board member for Bayer Healthcare; Dr. Lopez-Sendon, receiving grants from Bristol-Myers Squibb, Boehringer Ingelheim, and Bayer and consulting fees from Boehringer Ingelheim; Dr. Parkhomenko, receiving consulting

fees and grants from Borshchiagovsky Chemical-Pharmaceutical Plant; Dr. Verheugt, receiving lecture fees from Bayer and AstraZeneca and consulting fees from Bayer and Daiichi Sankyo; and Dr. Wallentin, receiving grants from Bristol-Myers Squibb, Pfizer, AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Schering-Plough, and Merck, consulting fees from Regado Biotechnologies, Portola, CSL Behring, Athera Biotechnologies, Boehringer Ingelheim, AstraZeneca, and GlaxoSmithKline, and lecture fees from Bristol-Myers Squibb, Pfizer, AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Schering-Plough, and Merck. No other potential conflict of interest relevant to this article was reported.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

APPENDIX

The authors' affiliations are as follows: the Duke Clinical Research Institute, Duke University Medical Center, Durham, NC (C.B.G., J.H.A., R.D.L., H.R.A.-K.); University of Glasgow, Glasgow, United Kingdom (J.J.V.M.); Boston University Medical Center, Boston (E.M.H.); Bristol-Myers Squibb, Princeton, NJ (M.H., M.G., P.M.); Lenox Hill Hospital, New York (J.A.); Oslo University Hospital, Oslo (D.A.); Dante Pazzanese Institute of Cardiology, São Paulo (A.A.); ECLA Estudios Cardiológicos Latinoamérica, Rosario, Argentina (M.C.B., R.D.); University of California San Francisco, San Francisco (J.D.E.); University of Alberta, Edmonton, Canada (J.A.E.); University of Missouri Health Care, Columbia (G.F.); University of New Mexico, Albuquerque (D.G.); Mayo Clinic, Rochester, MN (B.J.G.); Russian Cardiology Research Center, Moscow (S. Golitsyn); Tokai University School of Medicine, Kanagawa, Japan (S. Goto); Instituto Nacional de Cardiologia, Tlalpan CP, Mexico (A.G.H.); J.W. Goethe-University, Frankfurt, Germany (S.H.H.); University of Adelaide, Adelaide, Australia (J.H.); Motol University Hospital, Prague, Czech Republic (P.J.); Lady Davis Carmel Medical Center, Haifa, Israel (B.S.L.); Hospital Universitario La Paz, Madrid (J.L.L.-S.); St. John's Medical College, Bangalore, India (P.P.); Institute of Cardiology, Kiev, Ukraine (A.P.); Onze Lieve Vrouwe Gasthuis, Amsterdam (F.W.A.V.); Fuwai Hospital, Beijing (J.Z.); and Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden (L.W.).

REFERENCES

1. Hart RG, Benavente O, McBride R, Pearce LA. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. Ann Intern Med 1999;131:492-501.
2. ACTIVE Writing Group of the ACTIVE Investigators, Connolly S, Pogue J, et al. Clopidogrel plus aspirin versus oral anticoagulation for atrial fibrillation in the Atrial fibrillation Clopidogrel Trial with Irbesartan for prevention of Vascular Events (ACTIVE W): a randomised controlled trial. Lancet 2006;367:1903-12.
3. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med 2009;361:1139-51. [Erratum, N Engl J Med 2010;363:1877.]
4. Patel MR, Mahaffey KW, Garg J, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. N Engl J Med 2011. DOI: 10.1056/NEJMoa1009638.
5. Raghavan N, Frost CE, Yu Z, et al. Apixaban metabolism and pharmacokinetics after oral administration to humans. Drug Metab Dispos 2009;37:74-81.
6. Connolly SJ, Eikelboom J, Joyner C, et al. Apixaban in patients with atrial fibrillation. N Engl J Med 2011;364:806-17.
7. Lopes RD, Alexander JH, Al-Khatib SM, et al. Apixaban for reduction in stroke and other ThromboembolIc events in atrial fibrillation (ARISTOTLE) trial: design and rationale. Am Heart J 2010; 159:331-9. [Erratum, Am Heart J 2010;159: 1162.]
8. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briët E. A method to determine the optimal intensity of oral anticoagulant therapy. Thromb Haemost 1993; 69:236-9.
9. Schulman S, Kearon C, Subcommittee on Control of Anticoagulation of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in nonsurgical patients. J Thromb Haemost 2005; 3:692-4.
10. Jackson K, Gersh BJ, Stockbridge N, et al. Antithrombotic drug development for atrial fibrillation: proceedings, Washington, DC, July 25–27, 2005. Am Heart J 2008;155:829-40.
11. Go AS, Hylek EM, Borowsky LH, Phillips KA, Selby JV, Singer DE. Warfarin use among ambulatory patients with nonvalvular atrial fibrillation: the Anticoagulation and Risk Factors in Atrial Fibrillation (ATRIA) study. Ann Intern Med 1999;131: 927-34.
12. Alexander JH, Lopes RD, James S, et al. Apixaban with antiplatelet therapy after acute coronary syndrome. N Engl J Med 2011;365:699-708.
13. Eikelboom JW, Wallentin L, Connolly SJ, et al. Risk of bleeding with 2 doses of dabigatran compared with warfarin in older and younger patients with atrial fibrillation: an analysis of the Randomized Evaluation of Long-Term Anticoagulant Therapy (RE-LY) trial. Circulation 2011; 123:2363-72.
14. Frost C, Song Y, Barrett YC, Wang J, Li T, LaCreta F. Apixaban for prevention of acute ischemic events in patients with acute coronary syndromes. Presented at the XXIII Congress of the International Society on Thrombosis and Haemostasis, Kyoto, Japan, July 23–28, 2011. poster.
15. Mackman N. The role of tissue factor and factor VIIa in hemostasis. Anesth Analg 2009;108:1447-52.

Patient Education and Counseling 83 (2011) 252–255

Contents lists available at ScienceDirect

Patient Education and Counseling

journal homepage: www.elsevier.com/locate/pateducou

Educational/Counseling Model Health Care

Development of the Hepatitis C Self-Management Program

Erik J. Groessl [a,b,*], Kimberly R. Weingart [a], Allen L. Gifford [c,d], Steven M. Asch [e,f], Samuel B. Ho [a,b]

[a] VA San Diego Healthcare System, USA
[b] University of California San Diego, USA
[c] Edith Nourse Rogers Memorial Veterans Hospital, USA
[d] Boston University School of Public Health, USA
[e] VA Greater Los Angeles Healthcare System, USA
[f] Geffen School of Medicine at the University of California Los Angeles, USA

ARTICLE INFO

Article history:
Received 9 September 2009
Received in revised form 3 June 2010
Accepted 5 June 2010

Keywords:
Patient self-management
Cognitive/behavioral interventions
Hepatitis C

ABSTRACT

Objective: Chronic hepatitis C infection (HCV) is a major health problem that disproportionately affects people with limited resources. Many people with HCV are ineligible or refuse antiviral treatment, but less curative treatment options exist. These options include adhering to follow-up health visits, lifestyle changes, and avoiding hepatotoxins like alcohol. Herein, we describe a recently developed self-management program designed to assist HCV-infected patients with adherence and improve their health-related quality of life (HRQOL).

Methods: The development of the Hepatitis C Self-Management Program (HCV-SMP) was informed by scientific literature, qualitative interviews with HCV-infected patients, self-management training, and feedback from HCV clinical experts.

Results: The Hepatitis C Self-Management Program (HCV-SMP) is a multi-faceted program that employs cognitive-behavioral principles and is designed to provide HCV-infected people with knowledge and skills for improving their HRQOL. The program consists of six 2-h workshop sessions which are held weekly. The sessions consist of a variety of group activities, including disease-specific information dissemination, action planning, and problem-solving.

Conclusion: The intervention teaches skills for adhering to challenging treatment recommendations using a validated theoretical model. A randomized trial will test the efficacy of this novel HCV self-management program for improving HRQOL in a difficult to reach population.

Published by Elsevier Ireland Ltd.

1. Introduction

Hepatitis C virus (HCV) infects millions of people in the US and in Europe [1], and often co-occurs with substance abuse, homelessness, and impoverishment [2–5]. Long-term medical consequences of HCV include cirrhosis, hepatocellular carcinoma [6], and liver transplant [7]. HCV-infected individuals also experience physical and psychological symptoms and functional limitations [8,9]. Antiviral treatment can eliminate the virus in some patients [10,11], but the majority of patients are either ineligible for treatment, refuse treatment, or fail treatment [12].

Treatment recommendations for HCV-infected patients often include attending follow-up appointments, obtaining laboratory tests, undergoing psychiatric evaluation [13], abstention from alcohol [14,15], avoiding transmission of the virus to others, avoiding certain foods or medications, and losing weight [16–19]. Yet this population lacks resources in general, and many may not have the information or skills required to adhere to these recommendations successfully [8].

Like other chronic illnesses, patients can use social, behavioral, and cognitive skills to participate more effectively in caring for their HCV. These strategies may help patients self-manage symptoms, increase functionality, make more informed decisions about treatment, and increase their chances of completing treatment successfully. Chronic disease self-management programs are primarily grounded in social cognitive learning theory [20,21] and augment traditional patient education with behavioral change and problem-solving skills. Patients learn to identify and prioritize their problems, take appropriate action, and to solve problems in collaboration with health care professionals and supportive others.

Self-management programs for other chronic diseases have been studied extensively and have produced good results for

* Corresponding author at: Health Services Research and Development, VA San Diego Healthcare System, 3350 La Jolla Village Dr. 111 N-1, San Diego, CA 92161, USA. Tel.: +1 858 642 6347; fax: +1 858 552 4321.
E-mail address: egroessl@ucsd.edu (E.J. Groessl).

0738-3991/$ – see front matter. Published by Elsevier Ireland Ltd.
doi:10.1016/j.pec.2010.06.006

people with asthma, diabetes, heart failure, and arthritis among others [33–42]. Despite the clear need, these principles have not been applied to HCV previously. Our purpose is to describe the development of a self-management program adapted specifically for people with HCV.

2. Methods

2.1. Program development

The program was adapted for HCV-infected individuals from the Chronic Disease Self-Management Program (CDSMP) which is based primarily on cognitive-behavioral principles [22]. The adaptation/development of the HCV-SMP was informed by CDSMP training sessions, scientific literature review, qualitative interviews with HCV-infected patients [8], and by feedback from expert HCV clinicians and researchers. First, the developer of the HCV-SMP and staff members attended a week-long training session at the Stanford Patient Education Research Center [23] to fully understand the content and processes of the CDSMP. Going forward, Stanford Patient Education staff made themselves available for ongoing consultation.

Because chronic disease self-management approaches are often multi-faceted and allow participants to choose what they would like to change in their lives, generic health-related quality of life was chosen as the target outcome for the program. Accordingly, the main theoretical model used to inform the development of the program was the General Health Policy Model [24] which identifies symptoms (physical and psychological) and functioning as the primary determinants of well-being. Behaviors and cognitions shown to directly or indirectly impact HRQOL were considered for the program.

Scientific literature on the impact of HCV on HRQOL was also reviewed to further inform the development of the HCV-SMP and the main results of this review have been published [9]. The results identified fatigue [25], sexual dissatisfaction [26], depression, and anxiety [4] as some of the symptoms experienced by HCV-infected persons that could be addressed in a cognitive-behaviorally based program. Lifestyle factors associated with reduced HRQOL among persons with HCV included cigarette smoking [27] and current IV drug use. Other lifestyle factors known to impact the liver and/or HRQOL are alcohol use [14] and excess bodyweight [17,19]. Finally, there is evidence that HCV-infected patients report communication problems with physicians [28].

Qualitative interviews were conducted with 23 HCV-infected VA patients to learn more about the challenges of, and beneficial strategies for, living with HCV. The interviews were designed for the purpose of informing the development of an HCV-specific self-management program. Interview data were analyzed for themes by two independent raters using an HRQOL conceptual framework. More detail on the methods and results is available [8]. We found that: many participants had misconceptions about HCV, being diagnosed with HCV caused ongoing health concerns, HCV status affects interpersonal relationships, and the connection between substance use and HCV affects patient motivation for HCV treatment and substance abuse recovery.

After proposed changes to the existing CDSMP manual were developed, expert HCV clinicians reviewed these changes and provided feedback. Feedback included keeping the program to no more than six sessions, not intentionally emphasizing the common experience of veterans so that the program would generalize to others with HCV, limiting the time spent addressing severe active alcoholism or IV drug use primarily to referral, and devoting extra time to HCV treatment decision-making.

3. Results

The generic CDSMP adequately addressed many of the topics identified. Excess bodyweight was addressed in two modules on exercise and one module on healthy eating. Depression and anxiety were addressed in modules on cognitive symptom management, dealing with difficult emotions, depression management, self-talk, and multiple relaxation/stress reduction modules, so the only change was to spend more time on these areas. Interpersonal relationships were primarily discussed in the communication skills module, but also emerge during action planning and problem-solving. Garnering support from others is a tool for succeeding at behavioral change. We decided not to devote a module specifically to sexual satisfaction, but expected it to arise during the communication skills module where talking to others about HCV is discussed. Finally, patient-provider communication is addressed in modules on physician recommendations/medication usage, and informing the health care team.

To address a lack of accurate information about hepatitis C, we added HCV-specific education modules developed by the VA National Hepatitis C Office in sessions 1 and 2 [29]. Alcohol and drug usage in relation to HCV are addressed in the education modules, and an additional 20–25 min module was added that referred those with severe substance use to appropriate services, while providing tools for abstinence or harm reduction for those with minimal to moderate substance usage. We added material to the existing CDSMP module on fatigue, expanding it in length slightly. Another major addition to the existing program was adding a 40-min discussion panel about antiviral treatment in which 2–3 peers and an HCV health care provider described their experiences with antiviral treatment. This was followed by a module on shared treatment decisions which replaced a module on "making informed treatment decisions", with the former focusing on patient participation while the latter had focused on critically evaluating evidence as a consumer.

To keep the program to six sessions overall as recommended by the expert clinicians and researchers, we needed to eliminate a few modules. Namely, we eliminated modules on distraction from pain and advance directives for health care. Both topics seemed potentially useful to people with HCV, but not a priority based on our development process.

Two pilot programs were conducted to test the feasibility and participant reactions to the program. Although 12 patients attended the first assessment, 3 left before the program began, 9 attended the second session, 1 brought a friend who also attended, and eight attended regularly and completed baseline and post-intervention assessments. Overall patients showed improvement as shown in Table 1. From the pilot programs, we learned that participants usually attend the program regularly once they experienced the first session, and that conducting a larger study was feasible among VA patients.

The Hepatitis C Self-Management Program (HCV-SMP) is designed to provide HCV-infected people with the knowledge and skills they need to improve their current and future HRQOL. The HCV-SMP consists of six 2-h workshop sessions which are held weekly. The sessions consist of a variety of group activities, including disease-specific information dissemination, action planning, and problem-solving (see Table 2). Sessions are designed to promote participant interaction and the development of social support throughout. Program activities are modeled for patients by the co-leaders. Participants are provided with materials including an HCV information packet and resource guide, a copy of the book "Living a Healthy Life with Chronic Conditions" [30] and an audio recording of two relaxation exercises. The HCV-SMP is co-led by a health professional and a peer-leader. Both co-leaders are trained by studying the intervention manual and attending 40 h of training

Table 1
Pilot program outcome results.

Measures ($n = 8$)	Baseline mean (SD)	Follow-up mean (SD)	Effect size (d)	p-value (2-tail)
HRQOL—QWB score	.459 (.134)	.480 (.131)	.16	.318
Self-efficacy	5.69 (2.02)	6.73 (1.68)	.56	.029
Energy/fatigue	1.93 (1.38)	1.70 (.99)	.19	.501
Depression (CES-D 10)	18.6 (6.0)	15.4 (6.1)	.53	.344
HCV knowledge score	8.6 (2.7)	11.1 (1.6)	1.16	.038
Health distress	2.00 (1.22)	1.71 (1.41)	.22	.547
Exercise—self-report	4.4 (2.3)	5.0 (1.2)	.35	.553

at the Stanford Patient Education Research Center. A few of the most important modules from the CDSMP and new additions are described below.

Action plans are weekly goal setting activities in which goals are broken down into specific, manageable behaviors. The program leaders model this activity for participants by choosing an action plan and operationalizing it. Participants take turns stating their action plan and describing *What* they will do (something desirable and reasonable), *How Much* (i.e. time, effort), *How Often* (# of times week) and *When* (i.e. mornings, weekdays). Participants report their confidence level for each action plan on a scale from 0 to 100%. When participants are less than 70% sure they can complete their action plan, they modify their plan to make it more achievable.

Feedback/problem-solving sessions. Participants state their action plan and report their success completing the plan. If they were not successful, three problem-solving steps (identify the problem, list ideas to solve the problem, select one idea to try) are discussed with the group. The discussion follows a "brainstorm" format in which other group members suggest many ideas for solving the problem. The participant with the problem then chooses ideas to implement.

Substance use and HCV. Participants are directed to the VA web site for hepatitis and alcohol (www.hepatitis.va.gov/vahep). Participants are told about resources within the VA such as the Alcohol Drug Treatment Program (ADTP), and encouraged to seek help if they are surpassing VA guidelines for moderate drinking or actively using drugs. Many participants have been through substance use treatment or are involved with substance abuse recovery. The fit between abstinence/sobriety and self-management of HCV, barriers to sobriety, the pros/cons of abstinence, and strategies for reducing and abstaining from substance use are discussed.

Treatment discussion panel. The panel typically consists of 2–3 peers who have undergone antiviral treatment for HCV and one HCV health care provider. The peers describe their experience with HCV antiviral treatment and the health care provider offers additional insights based on clinical experience. Participants then ask questions about antiviral treatment. After the panel concludes, participants discuss active participation on treatment decision-making in light of the panel experience.

Table 2
Hepatitis C Self-Management Program modules.

	Modules
Session 1	Introductions—identifying common problems (20 min) Workshop overview and responsibilities (8–10 min) Hepatitis C—specific information with questions (25–30 min) Differences between acute and chronic conditions (12–15 min) Introduction to cognitive symptom management (10–12 min) Introduction to action plans (30 min) Brief review of topics and closing (10 min)
Session 2	Feedback/problem-solving session (20 min) Dealing with anger, fear and frustration (20 min) Introduction to exercise (30 min) Hepatitis C—specific information with questions (35 min) Making an action plan (20 min) Brief review of topics and closing (5 min)
Session 3	Feedback/problem-solving session (15–20 min) Fatigue management (15–20 min) Alcohol/drug use and hepatitis C (20–25 min) Better breathing (15 min) Endurance exercise (20 min) Making an action plan (15 min) Brief review of topics and closing (5 min)
Session 4	Feedback/problem-solving/making an action plan (20–25 min) Panel of veterans discussing interferon treatment (40 min) Making shared treatment decisions (10–15 min) Communication skills (15 min) Problem-solving (20 min) Muscle relaxation (10 min) Brief review of topics and closing (5 min)
Session 5	Feedback/problem-solving/making an action plan (20 min) Physician recommendations/medication usage (15–20) Healthy eating (25 min) Depression management (15 min) Self-talk (25 min) Guided imagery (20 min) Brief review of topics and closing (5 min)
Session 6	Feedback/problem-solving (20 min) Informing the health care team (10 min) Working with your health care professional (20 min) Looking back and planning for the future (35 min) Brief review of topics and closing (10 min)

4. Discussion and conclusion

Chronic hepatitis C infection is a major health problem both globally and in the US, that affects people who often lack resources [31,32]. Many of those infected may not be eligible or ready for antiviral treatment, and they receive treatment recommendations that are mostly behavioral in nature [12]. However, these recommendations are difficult to follow without additional support. Self-management programs are a novel option for helping HCV-infected patients adhere to treatment recommendations and improve their HRQOL. Similar programs for other chronic diseases have produced well-established results [33–42].

The HCV-SMP intervention was adapted specifically for people with HCV from an established chronic disease self-management program. The program teaches broad self-management skills that can be utilized to improve multiple aspects of health. The program's development was guided by the General Health Policy Model, and informed by scientific literature, qualitative interviews with HCV-infected patients, self-management training, and feedback from HCV clinical experts.

A possible limitation of the intervention is that it is not designed for people who are currently receiving antiviral treatment, but it does educate participants about antiviral treatment, assist them with treatment decision-making, and prepare them to succeed at treatment in the future. Although individuals on antiviral treatment may benefit from a self-management approach, such a program would likely focus more on side effects of treatment and adherence to antiviral medications. However, the current program can be adapted to help patients on antiviral treatment in the future.

The program is currently being evaluated among HCV-infected VA patients in a randomized controlled trial fund by VA HSR&D.

The study compares the HCV-SMP to an information-only comparison group. The study assesses self-reported outcomes such as HRQOL, energy, mental health, self-efficacy, HCV knowledge, and health behaviors in addition to clinical outcomes such as appointment attendance, antiviral treatment rates, and rates of sustained viral response. If efficacy is established in the RCT, the program will be implemented on a larger scale.

Conflict of interest

The authors do not have any conflicts of interest relating to this study or manuscript.

Acknowledgements

This study is funded by a research grant from the VA HSR&D, Project # IAC 05-067.

References

[1] Alter MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, Gao F, Moyer LA, et al. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. New Engl J Med 1999;341:556–62.
[2] Dominitz JA, Boyko EJ, Koepsell TD, Heagerty PJ, Maynard C, Sporleder JL, et al. Elevated prevalence of hepatitis C infection in users of United States veterans medical centers. Hepatology 2005;41:88–96.
[3] Gjeruldsen S, Loge JH, Myrvang B, Opjordsmoen S. Drug addiction in hepatitis C patients leads to a lower quality of life. Nord J Psychiatry 2006;60:157–61.
[4] Rowan PJ, Al-Jurdi R, Tavakoli-Tabasi S, Kunik ME, Satrom SL, El-Serag HB. Physical and psychosocial contributors to quality of life in veterans with hepatitis C not on antiviral therapy. J Clin Gastroenterol 2005;39:731–6.
[5] Rowan PJ, Tabasi S, Abdul-Latif M, Kunik ME, El-Serag HB. Psychosocial factors are the most common contraindications for antiviral therapy at initial evaluation in veterans with chronic hepatitis C. J Clin Gastroenterol 2004;38:530–4.
[6] Heintges T, Wands JR. Hepatitis C virus: epidemiology and transmission. Hepatology 1997;26:521–6.
[7] CDC. Center for Disease Control: recommendations for prevention and control of HCV infection and HCV-related chronic disease. MMWR 1998;47:1–10.
[8] Groessl EJ, Weingart KR, Kaplan RM, Clark JA, Gifford AL, Ho SB. Living with hepatitis C: qualitative interviews with hepatitis c-infected veterans. J Gen Intern Med 2008.
[9] Groessl EJ, Weingart KR, Kaplan RM, Ho SB. Health-related quality of life in HCV infected patients. Curr Hepatitis Rep 2007;6:169–75.
[10] Fried MW, Hadziyannis SJ. Treatment of chronic hepatitis C infection with peginterferons plus ribavirin. Semin Liver Dis 2004;24(Suppl. 2):47–54.
[11] Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales Jr FL, et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. New Engl J Med 2002;347:975–82.
[12] Groom H, Dieperink E, Nelson DB, Garrard J, Johnson JR, Ewing SL, et al. Outcomes of a Hepatitis C screening program at a large urban VA medical center. J Clin Gastroenterol 2008;42:97–106.
[13] VA, Hepatitis C Resource Center Program and National Hepatitis C Program Office. Management of psychiatric and substance use disorders in patients with hepatitis C: a reference for hepatitis C care providers; 2005. p. 1–36.
[14] Monto A, Patel K, Bostrom A, Pianko S, Pockros P, McHutchison JG, et al. Risks of a range of alcohol intake on hepatitis C-related fibrosis. Hepatology 2004;39:826–34.
[15] Pessione F, Degos F, Marcellin P, Duchatelle V, Njapoum C, Martinot-Peignoux M, et al. Effect of alcohol consumption on serum hepatitis C virus RNA and histological lesions in chronic hepatitis C. Hepatology 1998;27:1717–22.
[16] Adinolfi LE, Gambardella M, Andreana A, Tripodi MF, Utili R, Ruggiero G. Steatosis accelerates the progression of liver damage of chronic hepatitis C patients and correlates with specific HCV genotype and visceral obesity. Hepatology 2001;33:1358–64.
[17] Giannini E, Ceppa P, Testa R. Steatosis in chronic hepatitis C: can weight reduction improve therapeutic efficacy? J Hepatol 2001;35:432–3.
[18] Hickman IJ, Clouston AD, Macdonald GA, Purdie DM, Prins JB, Ash S, et al. Effect of weight reduction on liver histology and biochemistry in patients with chronic hepatitis C. Gut 2002;51:89–94.
[19] Hickman IJ, Jonsson JR, Prins JB, Ash S, Purdie DM, Clouston AD, et al. Modest weight loss and physical activity in overweight patients with chronic liver disease results in sustained improvements in alanine aminotransferase, fasting insulin, and quality of life. Gut 2004;53:413–9.
[20] Bandura A. Social foundations of thought and action. A social cognitive theory. Englewood Cliffs, NJ: Prentice Hall; 1986.
[21] Tobin D, Reynolds R, Holroyd K, Creer T. Self-management and social learning theory. In: Holroyd KA, Creer TL, editors. Self-management of chronic disease: handbook of clinical interventions and research. Orlando, FL: Academic Press; 1986 .
[22] Lorig K, Gonzalez VM, Laurent D. The chronic disease self-management workshop leaders manual. Stanford Patient Education Research Center; 1999
[23] Stanford Patient Education Research Center; 2010.
[24] Kaplan R, Anderson J. The general health policy model: an integrated approach. In: Spilker B, editor. Quality of life and pharmacoeconomics in clinical trials. Philadelphia: Lippincott-Raven; 1996. p. 309–22.
[25] Kramer L, Hofer H, Bauer E, Funk G, Formann E, Steindl-Munda P, et al. Relative impact of fatigue and subclinical cognitive brain dysfunction on health-related quality of life in chronic hepatitis C infection. AIDS 2005;19(Suppl. 3):S85–92.
[26] Danoff A, Khan O, Wan DW, Hurst L, Cohen D, Tenner CT, et al. Sexual dysfunction is highly prevalent among men with chronic hepatitis C virus infection and negatively impacts health-related quality of life. Amer J Gastroenterol 2006;101:1235–43.
[27] Balfour L, Cooper C, Kowal J, Tasca GA, Silverman A, Kane M, et al. Depression and cigarette smoking independently relate to reduced health-related quality of life among Canadians living with hepatitis C. Can J Gastroenterol 2006;20:81–6.
[28] Zickmund S, Hillis SL, Barnett MJ, Ippolito L, LaBrecque DR. Hepatitis C virus-infected patients report communication problems with physicians. Hepatology 2004;39:999–1007.
[29] HCRC. H.C.R.C., the hepatitis C education class. VA Healthcare System; 2006.
[30] Lorig K, Holman H, Sobel D, Laurent D, Gonzalez V, Minor M, et al. Living a healthy life with chronic conditions. Palo Alto, CA: Bull Publishing Co.; 2000 .
[31] Mishra G, Sninsky C, Roswell R, Fitzwilliam S, Hyams KC. Risk factors for hepatitis C virus infection among patients receiving health care in a Department of Veterans Affairs hospital. Digest Dis Sci 2003;48:815–20.
[32] Sloan KL, Straits-Tröster KA, Dominitz JA, Kivlahan DR. Hepatitis C tested prevalence and comorbidities among veterans in the US Northwest. J Clin Gastroenterol 2004;38:279–84.
[33] Barlow JH, Ellard DR. Psycho-educational interventions for children with chronic disease, parents and siblings: an overview of the research evidence base. Child Care Health Dev 2004;30:637–45.
[34] Barlow JH, Turner AP, Wright CC. Long-term outcomes of an arthritis self-management programme. Brit J Rheumatol 1998;37:1315–9.
[35] Bosmans J, de Bruijne M, van Hout H, van Marwijk H, Beekman A, Bouter L, et al. Cost-effectiveness of a disease management program for major depression in elderly primary care patients. J Gen Intern Med 2006;21:1020–6.
[36] Bourbeau J, Julien M, Maltais F, Rouleau M, Beaupre A, Begin R, et al. Reduction of hospital utilization in patients with chronic obstructive pulmonary disease: a disease-specific self-management intervention. Arch Intern Med 2003;163:585–91.
[37] Damush TM, Weinberger M, Perkins SM, Rao JK, Tierney WM, Qi R, et al. Randomized trial of a self-management program for primary care patients with acute low back pain: short-term effects. Arthritis Rheum 2003;49:179–86.
[38] de Asis ML, Greene R. A cost-effectiveness analysis of a peak flow-based asthma education and self-management plan in a high-cost population. J Asthma 2004;41:559–65.
[39] Latner JD. Self-help in the long-term treatment of obesity. Obes Rev 2001;2:87–97.
[40] Lorig K, Feigenbaum P, Regan C, Ung E, Holman HR. A comparison of lay-taught and professional-taught arthritis self-management courses. J Rheumatol 1986;13:763–7.
[41] Lorig KR, Ritter PL, Gonzalez VM. Hispanic chronic disease self-management: a randomized community-based outcome trial. Nurs Res 2003;52:361–9.
[42] Mårtensson J, Strömberg A, Dahlström U, Karlsson JE, Fridlund B. Patients with heart failure in primary health care: effects of a nurse-led intervention on health-related quality of life and depression. Eur J Heart Fail 2005;7:393–403.

JOURNAL OF VIRAL HEPATITIS
JVH

Journal of Viral Hepatitis, 2011, **18,** 358–368 doi:10.1111/j.1365-2893.2010.01328.x

The hepatitis C self-management programme: a randomized controlled trial

E. J. Groessl,[1,2] K. R. Weingart,[1] C. J. Stepnowsky,[1,2] A. L. Gifford,[3,4] S. M. Asch[5,6] and S. B. Ho[1,2] [1]*VA San Diego Healthcare System;* [2]*University of California San Diego, CA;* [3]*Edith Nourse Rogers Memorial Veterans Hospital, Bedford;* [4]*Boston University School of Public Health, Boston, MA;* [5]*VA Greater Los Angeles Healthcare System; and* [6]*David Geffen School of Medicine at the University of California, Los Angeles, CA, USA*

Received February 2010; accepted for publication March 2010

SUMMARY. Chronic hepatitis C (HCV) infection afflicts millions of people worldwide. While antiviral treatments are effective for some patients, many either cannot or choose not to receive antiviral treatment. Education about behavioural changes like alcohol avoidance and symptom management, in contrast, is universally recommended, particularly in HCV-infected persons from disadvantaged groups where liver risk factors are most prevalent. Self-management interventions are one option for fostering improved HCV knowledge and health-related quality of life (HRQOL). One hundred and thirty-two patients with VA with HCV (mean age of 54.6, 95% men, 41% ethnic minority, 83% unmarried, 72% unemployed/disabled, 48% homeless in last 5 years) were randomized to either a 6-week self-management workshop or an information-only intervention. The weekly 2-h self-management sessions were based on cognitive-behavioural principles and were adapted from an existing self-management programme that has been efficacious with other chronic diseases. HCV-specific modules were added. Outcomes including HRQOL, HCV knowledge, self-efficacy, depression, energy and health distress were measured at baseline and 6 weeks later. Data were analysed using ANOVA. When compared to the information-only group, participants attending the self-management workshop improved more on HCV knowledge ($P < 0.001$), HCV self-efficacy ($P = 0.011$), and SF-36 energy/vitality ($P = 0.040$). Similar trends were found for SF-36 physical functioning ($P = 0.055$) and health distress ($P = 0.055$). Attending the self-management programme improved disease knowledge and HRQOL 6 weeks later in this disadvantaged population. The intervention can improve the health of people with hepatitis C, independent of antiviral therapy. Future research will study longer-term outcomes, effects on antiviral treatment and costs.

Keywords: behavioural interventions, health-related quality of life, hepatitis C, self-management, US veterans.

Hepatitis C virus (HCV) infects about 1.8% of the US population [1] and often co-occurs with substance abuse problems, homelessness and impoverishment [2–5]. Long-term medical consequences of HCV include cirrhosis, hepatocellular carcinoma [6] and/or the need for liver transplant [7]. In addition, most HCV-infected individuals experience a variety of physical and psychological symptoms, functional limitations and impaired quality of life as a result of having HCV and co-existing chronic health problems [8,9].

Treatment with antiviral medications eliminates the virus in many patients [10,11], but lower success rates have been found outside of clinical trials [12]. Nevertheless, the vast majority of patients are either ineligible for treatment, refuse treatment, fail treatment, or treated with watchful waiting [13]. In fact, it is estimated that only about 20% of patients with VA with HCV have ever initiated antiviral treatment [12]. Common reasons for not receiving treatment are ongoing substance abuse, psychiatric disorders [12] and poor attendance of clinic appointments. Thus, despite ongoing improvements in antiviral treatments, there are few treatment alternatives for people with HCV.

Treatment recommendations for HCV-infected patients often include attending regular follow-up visits, obtaining additional laboratory tests, undergoing psychiatric

Abbreviations: ADTP, alcohol and drug treatment programme; CDSMP, chronic disease self-management programme; HQLQ, hepatitis quality of life questionnaire; HRQOL, health-related quality of life; QWB-SA, quality of well-being scale–self-administered; SVR, sustained virologic response; VAS, visual analogue type rating scale.

Correspondence: Erik J. Groessl, PhD, Health Services Research and Development, VA San Diego Healthcare System, 3350 La Jolla Village Dr. 111 N-1, San Diego, CA 92161, USA.
E-mail: egroessl@ucsd.edu

evaluation [14], abstention from alcohol [15,16], avoiding transmission of the virus, avoiding certain foods or medications, exercise/losing weight [17–20] and making decisions concerning antiviral treatment [21]. Yet, many patients may not have the information or skills required to adhere to these recommendations successfully [8].

As with other chronic illnesses, there are social, behavioural, and cognitive skills that patients can use to participate more effectively in the management of their HCV. These self-management strategies can help patients manage their symptoms, increase their functionality, make more informed decisions about treatment and potentially help prevent them from spreading the virus to others. Despite the need for such approaches, little or no research on self-management for hepatitis C has been undertaken.

Self-management interventions go beyond traditional patient education in that they are more comprehensive, focus more on facilitating change, teach problem-solving skills instead of primarily disseminating information [22] and engage the patient in the day-to-day management of his/her illness. Self-management programmes are primarily grounded in social cognitive learning theory [23,24]. Social learning theory emphasizes several dimensions along which change can occur, such as increasing self-efficacy and motivation, learning-specific illness management information and skills, enlisting support from a social network and monitoring symptoms and emotions [25]. Good self-management programmes typically address (i) disease management, medications, general health; (ii) role management and (iii) emotional management [26]. Patients learn to identify and prioritize their own problems, take appropriate action and enlist the support needed to solve these problems in collaboration with healthcare professionals and family.

The purpose of this study was to examine the effects of a hepatitis C self-management intervention on the quality of life of HCV-infected individuals who are not currently on or scheduled to start antiviral treatment.

METHODS

Study design

The study was a randomized controlled trial comparing the effects of the hepatitis C self-management programme to an HCV information-only intervention among VA patients with chronic hepatitis C. A total of 137 patients were recruited and randomized to one of the two intervention groups. After obtaining informed consent and a release of medical information, patients were randomized to one of two interventions and completed a baseline assessment. Self-management or comparison interventions were delivered over a 6-week period. Follow-up assessments occurred 6 weeks after the baseline assessment. The study was conducted at the VA San Diego Healthcare System, San Diego, CA, USA. Inclusion criteria were: (i) a confirmed diagnosis of chronic hepatitis C; (ii) US military veterans age 18 or older; (iii) eligible to receive care at the VA San Diego Healthcare System; (d) willing and able to attend six weekly programme sessions lasting about 2.5 h each. Exclusion criteria were: (i) ongoing antiviral treatment for hepatitis C or scheduled to start antiviral treatment within 6 months; (ii) residence in a geographical area outside of San Diego County (intervention attendance is too difficult); (iii) ongoing treatment for another life-threatening co-morbid illness; (iv) fatal co-morbidity (life expectancy <6 months indicated by treating physician).

Recruitment and retention

A multi-faceted recruitment approach was used. The approach was designed to reach a full spectrum of HCV-infected individuals, especially those who were impoverished and/or homeless, and often have been underrepresented in research. Patients were also told about the study by healthcare providers in the Hepatitis C Clinic and in primary care units at the VA San Diego Healthcare System. Flyers were posted in the Hepatitis C clinic, primary care clinics, the alcohol and drug treatment programme (ADTP) clinic, psychiatry and psychology clinics, common areas of the main VA hospital in San Diego, and at VA satellite clinics. Study recruiters also visited and established contacts at local recovery homes, homeless shelters, VA Stand Down San Diego and nonprofit organizations that serve veterans, such as St. Vincent De Paul and Veterans Village of San Diego (VVSD).

To minimize attrition and attendance problems, frequent contact with participants was maintained via phone and mail. The recruitment timeline for each cohort was approximately 1–2 months, so it was imperative to keep current phone numbers and addresses for interested persons. Prospective participants without working phone numbers were asked to contact study staff biweekly for study updates. After the initial baseline assessment, participants were given a printed schedule with the dates of their remaining assessments and/or self-management sessions.

Assignment and masking

As directed by referrals and flyers, potential participants called research project staff for more information on the study and eligibility screening. Participants who met study inclusion/exclusion criteria were added to a list of those interested in the study. Once a cohort of 20–25 participants had expressed interest, patients were randomized to one of the two interventions. Patients were recruited in nine cohorts over the course of 14 months. Research staff completed randomization for each participant using a computer-generated random number generator [27]. Blinding of participants and interventionists was not possible but both conditions could reasonably expect some improvement

in outcomes. Intervention staff were not blinded, but all assessments were self-report questionnaires.

Interventions

In an effort to standardize the amount of HCV-specific information available to participants and create a meaningful intervention for participants in the comparison group, both interventions were provided with an HCV-specific information packet upon completion of their baseline assessment. The packet contained information that is available to all patients with VA nationally through their healthcare providers and/or via the Internet and the Hepatitis C Resource Centers. In addition, all participants received a patient resource guide, containing an indexed list of health-related organizations and telephone numbers. Clinical care for chronic HCV infection continued as usual for participants in both interventions.

HCV information only

The HCV-specific information booklet consisted of printed materials that are currently available in the Hepatitis C Clinic and in other usual care clinics for patients with Hepatitis C. The materials included brochures and handouts titled 'Overview of Hepatitis C', 'Understanding the Side Effects of Interferon Therapy', Coping with Hepatitis C: Diet and Nutrition' and 'Talking with Others About Hepatitis C' to name a few. In general, the materials described hepatitis C and cover the topics of preventing the transmission of hepatitis C, avoiding certain foods and medications, discussing one's diagnosis with others, antiviral treatment and general behavioural recommendations. Also, included were printouts of the same presentation used for the self-management workshop intervention. The presentation was prepared and disseminated by the VA National Hepatitis C Resource Centers. Patients in the information-only intervention were able to ask their healthcare providers about the information they received at any time. Questions specific to the research study were directed to the project coordinator.

Hepatitis C self-management programme

Programme development

The programme was adapted for veterans with chronic HCV infection from the chronic disease self-management programme (CDSMP) developed by Lorig *et al.* [28]. The adaptation and development of the HCV-SMP were informed by a qualitative pilot study with HCV-infected veterans [8], scientific literature, CDSMP training sessions at Stanford University and input from collaborators at the VA Hepatitis C Resource Centers and the VA HIV/Hepatitis QUERI Coordinating Center. Our main modifications to the CDSMP include the addition of three HCV-specific education modules developed by the VA National Hepatitis C Resource Centers, each about 20 min in length, a panel discussion with patients who had already completed antiviral treatment, treatment decision-making, discussion of substance use disorders and treatment, and increased time devoted to psychological issues, evaluating alternative treatments and communication with healthcare providers. To keep the intervention to a maximum of six sessions similar to the CDSMP, modules on advanced directives and cognitive distraction were removed and other modules were shortened.

The hepatitis C self-management programme (HCV-SMP) is designed to provide HCV-infected people with the knowledge and skills they need to improve their current and future health and health-related quality of life (HRQOL). The programme is based primarily on cognitive-behavioural principles and is multi-faceted, focusing not only on hepatitis C, but also on general health improvement. The programme incorporates a client-centred, empowerment-based approach that allows the participants to prioritize which elements of the programme that they would like to focus on most.

The HCV-SMP consists of six weekly workshop sessions that are 2–2.5 h in length. The sessions consist of a variety of group activities, including disease-specific information dissemination, problem solving, development of action plans, and re-evaluation and revision of action plans. Action plans consist of small behavioural changes that each participant aims to accomplish during the week between sessions. Participants report back on their success and are assisted with problem solving when they are not successful. In addition to the interaction and the exchange of social support that takes place between participants before and after the sessions, a 20-min break occurs at the middle of each session, and often encourages additional interactions among participants. Patients are encouraged, but not required, to discuss their experiences and problems with other group members and with the group as a whole. All programme activities are modeled for patients by the group leaders, and participants can opt out of activities at any time. In addition to the HCV information packet and the patients' resources guide, each patient attending the HCV-SMP receives a copy of the book Living a Healthy Life with Chronic Conditions [29] and a cassette tape or CD containing two relaxation exercises. Light refreshments were served at each session because the sessions are up to 2.5 h long, and fatigue can be a problem for patients with hepatitis C.

The HCV-SMP was co-led by a health professional and a peer-leader. Research has shown that peer-leaders can teach self-management as well as health professionals when properly trained and that patients may obtain additional benefit from being able to identify with someone who has the same illness [30–32]. To maintain consistency and fidelity of the multiple intervention cohorts, all workshops were delivered by the same leader and peer leader.

Between each workshop session, participants were contacted by the group leader via phone. The phone call is

aimed to check on the progress of the participant's action plan and remind them of the upcoming group. After the first workshop session, participants are given the option to call each other once during the week instead of being called by the leader. Based on data from preliminary studies, participants look forward to the phone call, but many prefer to be called by the group leader.

Measures

The primary outcomes for the study were generic and disease-specific HRQOL. It was hypothesized that self-management participants would have greater improvements in HRQOL than information-only participants. Secondary outcomes were examined with the same hypothesis as the primary outcomes. These measures included HCV-related knowledge, self-efficacy, depression, energy/fatigue, alcohol use and health distress. A sociodemographic questionnaire was used to assess variables such as age, gender, race/ethnicity, education level, marital status, employment status, transportation and living situation.

Generic health-related quality of life

The SF-36 and the quality of well-being scale–self-administered (QWB-SA) were used to assess generic HRQOL. The SF-36 is a generic, descriptive measure of HRQOL that includes eight domains: physical functioning, role limitations – physical, bodily pain, general health perceptions, vitality, social functioning, role limitations – emotional, and overall mental health [33]. The measure has well-established norms, and the reliability and validity of the SF-36 are well documented [34–36]. The QWB-SA is a preference-based measure that produces a single total summary score that is appropriate for cost-effectiveness analysis [37–42]. The QWB-SA also asks respondents to rate their global health over the past 12 months using a visual analogue type rating scale (VAS), anchored at 0 and 100.

Hepatitis C-specific quality of life

The hepatitis quality of life questionnaire (HQLQ) is a hepatitis-specific quality of life measure administered in conjunction with the SF-36. Domains assessed include physical and social functional limitations, health distress, sleep and vitality/energy. It was validated in HCV-infected individuals [43] and has been used in a number of HCV clinical trials [44–46].

Hepatitis C knowledge

HCV knowledge was assessed using a measure developed by the study investigators to assess topics covered in the VA National Hepatitis C Resource Center presentation that is available to all veterans via the Internet or through their healthcare providers. The measure consists of 15 questions with Yes/No/Not Sure response options.

Self-efficacy

Confidence in the ability to perform HCV-related self-management activities was measured using HCV-specific self-efficacy questions adapted from other studies [47].

Energy/Fatigue

Energy/Fatigue was measured using a five-item scale adapted from the medical outcomes study [47]. The five items ask about how often symptoms of fatigue are present on average during the past month.

Depression

Depression was assessed using the Center for Epidemiologic Studies Short Depression Scale (CES-D 10) [48]. A score of 10 or greater is considered depressed. Normative data from assorted chronic illnesses are available [32].

Health distress

Health distress health-related anxiety is a common symptom in people with a chronic, potentially life-threatening illness. Health distress was measured with a four-item scale [47].

Alcohol usage

The Alcohol Usage Disorders Identification Test (AUDIT) was used to measure alcohol usage [49]. The measure is 10 questions and can be completed in 1–2 min. It is well validated and has been used extensively for many years [49–52].

Statistical analyses

Means and proportions of demographic and clinical characteristics for each group were compared with independent sample t-tests and chi-square analyses. Continuous variables were checked for violations of normality prior to t-test analysis. Equal variances were not assumed for the t-tests. Intention-to-treat analyses included repeated measures ANOVA and ANCOVA, which were used to compare differences between the two groups from baseline to the 6-week follow-up. Age and self-reported history of homelessness in the past 5 years were tested as covariates in the repeated measures ANOVA analyses. The covariates were retained in the model when significant at $P < 0.05$ level.

Based on previous studies of self-management programmes [53,54], the study was powered to have an 80% chance, with alpha of 0.05, of detecting a medium effect size of 0.25, and required 64 subjects per group or 128 total subjects after attrition.

RESULTS

Figure 1 shows the identification, enrolment and randomization of patients in the study. Of the 372 individuals who expressed interest in the study, 81 were ineligible because they were currently on, had successfully completed, or were scheduled to start antiviral treatment in the next 6 months

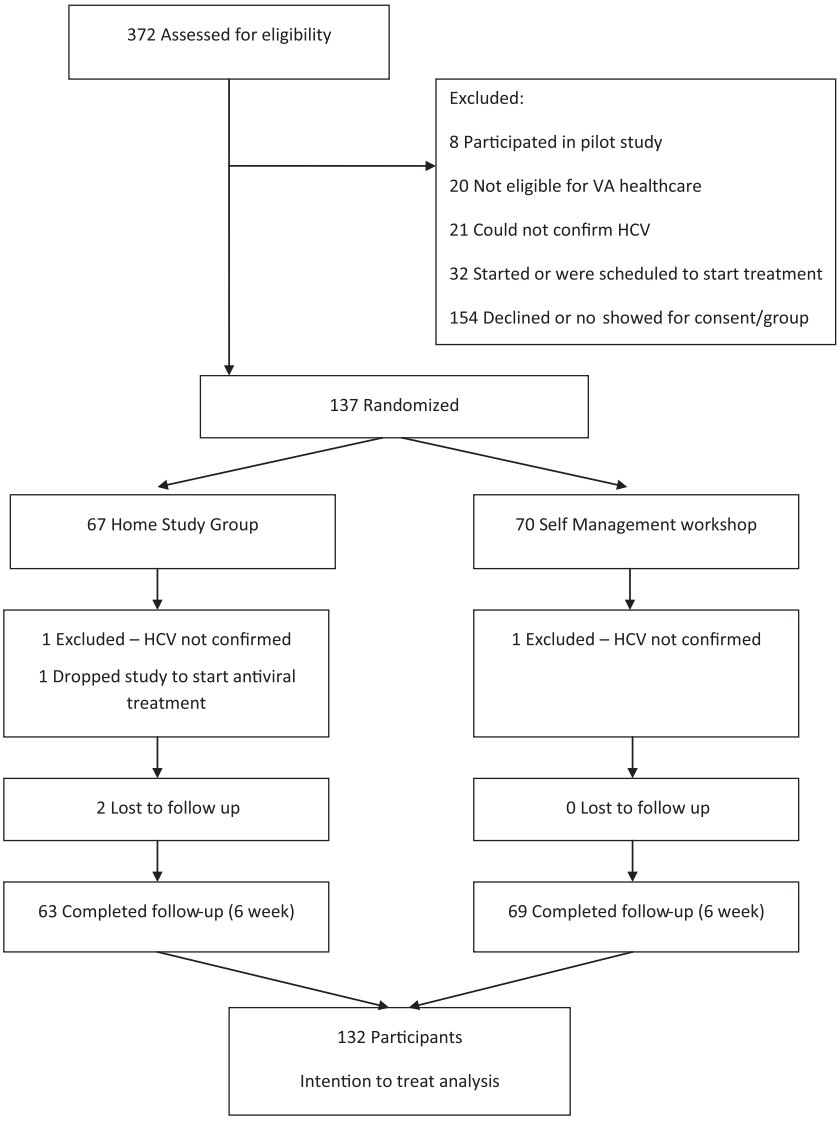


Fig. 1 Recruitment, randomization and retention flowchart.

(n = 32), were not eligible for VA healthcare (n = 20), their HCV could not be confirmed (n = 21), or had previously participated in the self-management pilot intervention (n = 8). An additional 154 individuals were initially eligible but declined enrolment when it became available. All 137 participants were recruited and consented between May 2007 and November 2008. Confirmation of HCV infection could not be produced from any source for two people, who were subsequently unenrolled. One other patient withdrew from the study and analysis. Of the remaining 134 participants, two information-only participants did not complete the 6-week follow-up assessment, leaving data on 132 participants for analysis.

Overall, the 132 participants had a mean age of 54.6, and 95% were men. Non-Hispanic white race/ethnicity was self-reported by 59% of the sample, with African Americans (24%) and Hispanics (10%) representing the next largest groups. The overall participant sample was 79% divorced, separated, or never married, 72% were disabled or unemployed, 35% never attended college, 68% relied on public transport or other people for transportation, 14% were homeless while another 36% lived in group living situations or with relatives, and 48% of the sample reported being homeless at some point in the last 5 years. The mean self-report year of contracting HCV was 1983, while mean self-report year of being diagnosed with HCV was 1997.

The two experimental groups were compared on three continuous and eight categorical variables. On average, participants in the information-only group were significantly older (t (1,130) = 3.07, P = 0.003) and more likely to be homeless in the past 5 years when compared with those randomized to the self-management programme. However, there was no difference between the two groups in the proportion that were currently homeless (See Table 1). Thus, age and self-reported history of homelessness were tested as covariates in the repeated measures ANOVA

Table 1 Participant demographics and characteristics

Variable	Information-only	Self-management workshop	*P*-value
	Mean (SD) or %		
Mean age (SD)	56.4 (7.2)	53.0 (5.2)	0.003**
Gender			
Male	95	96	0.909
Female	5	4	
Education %			
<high school grad	8	9	0.558
High school grad/GED	22	30	
Some college	57	45	
College graduate	13	16	
Ethnicity %			
African American	29	20	0.393
Asian/Pacific Islander	1	1	
Hispanic	8	12	
Native American	3	0	
Non-Hispanic White	57	61	
Other	2	6	
Employment %			
Unemployed	29	30	0.104
Disabled	37	49	
Retired	16	3	
Employed	14	15	
Other	5	3	
Marital status %			
Married	19	16	0.079
Widowed	2	4	
Divorced or separated	63	47	
Never married	16	33	
Residential %			
Homeless	13	15	0.577
Group living residence	30	35	
With relatives	3	4	
Apartment	33	36	
Own house	21	10	
Homeless in the past 5 years %	59	38	0.015*
Transportation %			
Drive own vehicle	40	26	0.241
Public transportation	51	64	
Other	9	10	
Mean years since contraction (SD) – self-report	25.1 (13.0)	21.6 (11.9)	0.152
Mean years since diagnosis (SD) – self-report	9.9 (8.9)	9.9 (9.6)	0.995

**$P < 0.01$, *$P < 0.05$.

analyses. The covariates were retained in the model when significant at $P < 0.05$ level.

Participants randomized to HCV-SMP attended an average of 5.2 of six sessions (87%). As shown in Table 2, participants in the self-management group had better outcomes that were statistically significant at $P < 0.05$ on a number of important variables. SF-36 energy/vitality scores increased almost 5 points in the self-management workshop while the mean scores in the information-only group decreased 2.6 points ($P = 0.040$). Statistically significant differences over time were also found for HCV knowledge ($F(1,129) = 20.35$, $P < 0.001$) and HCV-related self-efficacy ($F(1,130) = 6.57$, $P = 0.011$). Trends towards greater improvements for HCV-SMP participants were found for SF-36 physical functioning, SF-36 bodily pain, health distress, depression and VAS global health. Overall, self-management workshop participants had greater improvements on 18 of the 22 variables.

DISCUSSION

Antiviral medication can eliminate HCV for some infected patients, but depending on patient and viral characteristics and treatment compliance, viral clearance for some patient subgroups is very low [12]. Furthermore, patients are very often either ineligible for treatment, refuse treatment, or are offered watchful waiting [13]. Substantial improvements in treatment success rates are expected over the next couple of years [55], but there is a need for programmes that can help patients improve HRQOL while preparing them to become treatment candidates, and increasing their chances for achieving an SVR if they are eventually treated. Despite the need for such approaches, there has been little or no research conducted with self-management programmes for patients with hepatitis C. The randomized, controlled trial described here assesses the efficacy of the hepatitis C self-management programme for improving the HRQOL and other outcomes for patients with VA living with chronic hepatitis C.

The HCV self-management programme was well attended and produced significant improvements along a number of dimensions of HRQOL and other outcomes. Improving health-related knowledge is a main goal of most self-management programmes. While increased knowledge may improve healthcare satisfaction in consumers, it is not always considered sufficient for producing changes in other health outcomes. Improved self-efficacy (confidence in one's own ability to manage one's health condition) has been conceptually linked to changes in health outcomes in the social cognitive model on which the HCV self-management programme is based [23,24,56]. Self-efficacy is thought to interact with outcome expectations and sociostructural factors in the setting of goals and performance of behaviours. Behaviours in turn affect health outcomes, but it is also hypothesized that improved self-efficacy alone can improve quality of life, by reducing anxiety or distress associated with having a chronic disease and reducing powerlessness [57].

Table 2 Primary and secondary outcomes

Measure	Baseline	6-week follow-up	Change	*P*-value
HCV knowledge				
Home	8.6	9.9	1.3	<0.001**
Workshop	8.8	12.2	3.4	
Self-efficacy				
Home	7.18	7.09	−0.09	0.011*
Workshop	7.10	7.86	0.76	
Energy				
Home	2.46	2.61	0.15	0.464
Workshop	2.45	2.50	0.05	
CES-D				
Home	10.1	11.1	1.0	0.093†
Workshop	11.3	10.6	−0.7	
Health distress				
Home	2.09	2.14	0.05	0.055†
Workshop	2.14	1.87	−0.27	
QWB				
Home	0.503	0.517	0.014	0.255
Workshop	0.534	0.576	0.042	
Global health status – (VAS 0–100)				
Home	59.4	59.0	−0.4	0.105†
Workshop	58.4	63.9	5.5	
SF-36 physical functioning				
Home	66.9	63.3	−3.6	0.055†
Workshop	66.7	70.10	3.3	
SF-36 role physical				
Home	56.9	57.1	0.2	0.219
Workshop	57.1	62.9	5.8	
SF-36 bodily pain				
Home	43.5	51.3	7.8	0.073†
Workshop	49.4	50.5	0.9	
SF-36 general health				
Home	47.9	49.7	1.8	0.822
Workshop	51.5	52.6	1.1	
SF-36 vitality/energy				
Home	52.1	50.0	−2.1	0.040*
Workshop	46.4	51.0	4.6	
SF-36 social functioning				
Home	61.7	60.7	−1.0	0.687
Workshop	60.9	61.8	0.9	
SF-36 role emotional				
Home	64.2	63.6	−0.6	0.881
Workshop	64.4	64.6	0.2	
SF-36 mental health (covariate: age)				
Home	61.9	62.9	1.0	0.972
Workshop	60.5	61.6	1.1	
SF-36 health change				
Home	52.0	52.4	0.4	0.200
Workshop	53.6	59.1	5.5	
SF-36 PCS				
Home	40.8	41.3	0.5	0.375
Workshop	42.0	43.7	1.7	
SF-36 MCS				
Home	42.6	42.1	−0.5	0.583
Workshop	41.0	41.6	0.6	
HQLQ positive well being				
Home	53.9	55.2	1.3	0.772
Workshop	51.4	51.8	0.5	
HQLQ health distress (covariate: age)				
Home	62.0	58.7	−3.3	0.099†
Workshop	58.4	62.0	3.6	
HQLQ hepatitis-specific limitations				
Home	69.9	71.9	2.0	0.620
Workshop	75.2	75.0	−0.2	
HQLQ hepatitis-specific health distress (covariate: age)				
Home	68.1	65.4	−2.7	0.472
Workshop	71.2	71.5	0.3	

HQLQ, hepatitis quality of life questionnaire; VAS, visual analogue type rating scale. $^{**}P < 0.01$, $^{*}P < 0.05$, $^{\dagger}P < 0.10$.

The finding that the HCV self-management intervention improved scores on the energy/vitality subscale of the SF-36 is important. Fatigue is the most commonly reported symptoms of chronic HCV infection [58,59]. Fatigue has also been implicated as one of the symptoms most likely to impact other aspects of HRQOL, such as physical, occupational and/or social functioning.

Six other outcomes showed trends towards benefit from the HCV self-management intervention. Although we ran 22 separate statistical analyses, only one outcome would be expected to be significant by chance at the $P < 0.05$ level and 2 at a $P < 0.10$ level. Thus, there is fairly strong evidence that the intervention produced significant benefit for participants, but the study may have been underpowered to detect other effects not found to be significant. Given our final sample for these analyses, the study was powered to detect medium *F*-test effects (0.25) 81% of the time. Some of the effects were smaller than 0.25, but approached clinical significance. For example, a difference of 0.030 is considered clinically significant for the QWB-SA [60], and our study produced group differences of 0.028 after just 6 weeks. Thus, our study may have benefitted from a larger sample size.

Effect sizes in the range of 0.20–0.25 and adjusted mean differences between groups for the SF-36 fatigue and physical functioning subscales of 6.7 and 6.9 respectively are similar to those found in other studies of chronic disease self-management programmes [32,61,62]. Other self-management studies have found differences on other outcome measures such pain, disability, etc., but found no differences on validated measures of HRQOL [63–65]. In fact, most self-management studies do not formally measure HRQOL at all

[66]. This literature suggests that although the differences we have found are not large, they are meaningful, and it is important to start measuring self-management outcomes with widely used HRQOL measures such as the SF-36.

Chronic disease self-management interventions have also impacted other outcomes such as reduced health care utilization and increased health behaviours [54,67]. Our intervention was not designed to reduce health care costs and encourages HCV-infected patients to be properly evaluated for antiviral treatment which can be costly. However, if participants in the self-management programme are more likely to go on to be treated, there are well-established health benefits for those that achieve an SVR. Rates of antiviral treatment and response to treatment will be forthcoming. In addition, we plan to examine the costs of conducting the self-management intervention and conduct cost-effectiveness analyses. In general, we expect that the per person cost of the intervention will be low, because the intervention is delivered in group format for 6 weeks using existing facilities with few materials required. The cost analysis of a similar 6-week self-management programme found the intervention cost approximately £101 per person [68].

Although not all outcomes improved significantly, the results are notable for a few important reasons. First, the project was successful in recruiting and retaining vulnerable veterans with few resources who have faced many life challenges, such as those with a history of substance use, homelessness and psychological disorders. Overall, participants appear to be representative of HCV-infected individuals in VA care and in the general community [12]. Thus, the intervention may be beneficial to a large majority of HCV-infected individuals, but should be studied for efficacy among nonveterans.

Second, the current results were achieved between baseline and a 6-week follow-up. Although some variables are expected to change the most directly between pre and post intervention, HRQOL outcomes combine many factors and can be hard to change in the short term with behavioural and cognitively oriented interventions. Often, behavioural interventions must be sustained for a longer period of time to achieve significant changes [69,70].

One important unanswered question considers how many HCV-infected individuals will actually attend a 6-week self-management programme. This is difficult to determine because this is the first Hepatitis C self-management programme that has been developed. However, only 34 of 373 individuals who contacted study staff to inquire about the project declined because they did not want to attend. Many others expressed interest but could not attend at the time offered. A study of an asthma self-management programme examined the 'reach' of their programme and found that 474 of the 1303 in the target population with asthma actually enrolled in the programme [71]. This suggests that a sizable portion of the HCV-infected population who have not been successfully treated may be receptive to this type of intervention. However, more translational research is needed to more fully answer this question.

It is important to note that the intervention is not designed for patients who are receiving antiviral treatment. Although they comprise a minority of all HCV-infected patients, patients on medication must cope with a variety of difficult side effects and manage their medications carefully. Thus, they may need even more supportive resources because antiviral treatment poses unique challenges, but patients on antiviral treatment are also usually higher functioning on average than those who are ineligible for treatment. In general, more healthcare resources are devoted to helping patients receiving antiviral treatment, which highlights the need for interventions like the HCV-SMP, which serve those who are not on antiviral treatment. Conceivably, the current programme could be extended to patients on antiviral therapy, but efficacy would need to be tested with patients on antiviral treatment.

Study enrolment was limited to patients with VA with hepatitis C, so our conclusions may not generalize well beyond the VA system. However, similar self-management programmes addressing other diseases have been studied extensively in nonveteran samples and have regularly produced good results. The intervention will be studied in nonveteran samples in the future.

A possible limitation of the study is that eligibility for the study and the intervention itself required participants to attend six self-management sessions at a large medical centre. Classes were typically held on weekday afternoons because that was the time when most individuals were available. This excluded people who could not attend during these times, or those who lived in distant areas. Thus, it may be helpful to explore options for greater time flexibility and delivering this type of intervention to remote sites using medical informatics.

In conclusion, the hepatitis C self-management programme produced a number of important health benefits among patients with VA with chronic hepatitis C infection. There are few, if any, interventions available for HCV-infected individuals who have failed, or are not immediate candidates for antiviral therapy and this group makes up almost three quarters of all HCV-infected persons. The programme is broad and may impact general health and other co-morbid disorders instead of just HCV itself. Planned research will examine longer term outcomes and clinical outcomes and costs. If successful, the intervention is a model for larger scale implementation in the VA and other systems with large numbers of disadvantaged HCV-infected patients.

ACKNOWLEDGEMENTS

This study is funded by a research grant from the VA HSR&D, Project # IAC 05-067. This project is registered at clinical Trails.gov, registration # NCT00328042.

DISCLOSURE STATEMENT

The authors do not have any conflicts of interest relating to this study or manuscript.

REFERENCES

1 Alter MJ, Kruszon-Moran D, Nainan OV *et al.* The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med* 1999; 341: 556–562.
2 Dominitz JA, Boyko EJ, Koepsell TD *et al.* Elevated prevalence of hepatitis C infection in users of United States veterans medical centers. *Hepatology* 2005; 41(1): 88–96.
3 Gjeruldsen S, Loge JH, Myrvang B, Opjordsmoen S. Drug addiction in hepatitis C patients leads to a lower quality of life. *Nord J Psychiatry* 2006; 60(2): 157–161.
4 Rowan PJ, Al-Jurdi R, Tavakoli-Tabasi S, Kunik ME, Satrom SL, El-Serag HB. Physical and psychosocial contributors to quality of life in veterans with hepatitis C not on antiviral therapy. *J Clin Gastroenterol* 2005; 39(8): 731–736.
5 Rowan PJ, Tabasi S, Abdul-Latif M, Kunik ME, El-Serag HB. Psychosocial factors are the most common contraindications for antiviral therapy at initial evaluation in veterans with chronic hepatitis C. *J Clin Gastroenterol* 2004; 38(6): 530–534.
6 Heintges T, Wands JR. Hepatitis C virus: epidemiology and transmission. *Hepatology* 1997; 26(3): 521–526.
7 CDC. Center for Disease Control: recommendations for prevention and control of HCV infection and HCV-related chronic disease. *MMWR Morb. Mortal. Wkly. Rep.* 1998; 47(RR-19): 1–10.
8 Groessl EJ, Weingart KR, Kaplan RM, Clark JA, Gifford AL, Ho SB. Living with hepatitis C: qualitative interviews with hepatitis C-infected veterans. *J Gen Intern Med* 2008; 23(12): 1959–1965.
9 Groessl EJ, Weingart KR, Kaplan RM, Ho SB. Health-related quality of life in HCV infected patients. *Curr Hepat Rep* 2007; 6(4): 169–175.
10 Fried MW, Hadziyannis SJ. Treatment of chronic hepatitis C infection with peginterferons plus ribavirin. *Semin Liver Dis* 2004; 24(Suppl. 2): 47–54.
11 Fried MW, Shiffman ML, Reddy KR *et al.* Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002; 347(13): 975–982.
12 Ho SB, Groessl E, Dollarhide A, Robinson S, Kravetz D, Dieperink E. Management of chronic hepatitis C in veterans: the potential of integrated care models. *Am J Gastroenterol* 2008; 103(7): 1810–1823.
13 Groom H, Dieperink E, Nelson DB *et al.* Outcomes of a Hepatitis C screening program at a large urban VA medical center. *J Clin Gastroenterol* 2008; 42(1): 97–106.
14 VA Hepatitis C Resource Center Program and National Hepatitis C Program Office. Management of Psychiatric and Substance Use Disorders in Patients with Hepatitis C: A Reference for Hepatitis C Care Providers. Washington DC: Department of Veteran Affairs, 2005: 1–36.
15 Monto A, Patel K, Bostrom A *et al.* Risks of a range of alcohol intake on hepatitis C-related fibrosis. *Hepatology* 2004; 39(3): 826–834.
16 Pessione F, Degos F, Marcelliin P *et al.* Effect of alcohol consumption on serum hepatitis C virus RNA and histological lesions in chronic hepatitis C. *Hepatology* 1998; 27: 1717–1722.
17 Adinolfi LE, Gambardella M, Andreana A, Tripodi MF, Utili R, Ruggiero G. Steatosis accelerates the progression of liver damage of chronic hepatitis C patients and correlates with specific HCV genotype and visceral obesity. *Hepatology* 2001; 33(6): 1358–1364.
18 Giannini E, Ceppa P, Testa R. Steatosis in chronic hepatitis C: can weight reduction improve therapeutic efficacy? *J Hepatol* 2001; 35(3): 432–433.
19 Hickman IJ, Clouston AD, Macdonald GA *et al.* Effect of weight reduction on liver histology and biochemistry in patients with chronic hepatitis C. *Gut* 2002; 51(1): 89–94.
20 Hickman IJ, Jonsson JR, Prins JB *et al.* Modest weight loss and physical activity in overweight patients with chronic liver disease results in sustained improvements in alanine aminotransferase, fasting insulin, and quality of life. *Gut* 2004; 53(3): 413–419.
21 Yee HS, Currie SL, Darling JM, Wright TL Management and treatment of hepatitis C viral infection: recommendations from the department of veterans affairs hepatitis C resource center program and the national hepatitis C program office. *Am J Gastroenterol* 2006; 101(10): 2360–2378.
22 Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. *JAMA* 2002; 288(19): 2469–2475.
23 Bandura A. Social Foundations of Thought and Action. A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice Hall, 1986.
24 Tobin D, Reynolds R, Holroyd K, Creer T. Self-management and social learning theory. In: KA H, TL C, eds. Self-management of Chronic Disease: Handbook of Clinical Interventions and Research. Orlando, FL: Academic Press, 1986.
25 Von Korff M, Gruman J, Schaefer J, Curry SJ, Wagner EH. Collaborative management of chronic illness. *Ann Intern Med* 1997; 127(12): 1097–1102.
26 RWJF. Essential Elements of Self-management Interventions. Seattle, WA: The Robert Wood Johnson Foundation. The Center for the Advancement of Health, 2001.
27 Haahr M. Random.org: True random number service, 2004. Available at: http://www.random.org/nform.html (accessed 10 November 2008).
28 Lorig K, Gonzalez VM, Laurent D. The Chronic Disease Self-management Workshop Leaders Manual. Palo Alto, CA: Stanford Patient Education Research Center, 1999.
29 Lorig K, Homan H, Sobel D *et al.* Living a Healthy Life with Chronic Conditions. Palo Alto, CA: Bull Publishing Co., 2000.
30 Cohen JL, Sauter SV, deVellis RF, deVellis BM. Evaluation of arthritis self-management courses led by laypersons and by professionals. *Arthritis Rheum* 1986; 29(3): 388–393.
31 Lorig K, Feigenbaum P, Regan C, Ung E, Chastain RL, Holman HR. A comparison of lay-taught and professional-taught arthritis self-management courses. *J Rheumatol* 1986; 13(4): 763–767.

32 Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. *Eff Clin Pract* 2001; 4(6): 256–262.

33 Kosinski M, Keller SD, Hatoum HT, Kong SX, Ware Jr JE. The SF-36 Health Survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: tests of data quality, scaling assumptions and score reliability. *Med Care* 1999; 37(5 Suppl.): MS10–MS22.

34 Stewart AL, Hays RD, Ware Jr JE. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care* 1988; 26(7): 724–735.

35 Ware Jr JE, Kosinski M, Turner-Bowker DM, Gandek B. How to Score Version 2 of the SF-12 Health Survey (With a Supplement Documenting Version 1). Lincoln, RI: QualityMetric Incoporated, 2002.

36 Ware JE, Kosinski MA, Keller SD. SF-36 Physical and Mental Health Summary Scales: A User's Manual, 5th edn. Boston, MA: Health Assessment Lab, New England Medical Center, 1994.

37 Kaplan RM, Anderson JP. The quality of well-being scale: rationale for a single quality of life index. In: Walker SR, Rosser R, eds. Centre for Medicines Research Workshop CIBA Foundation, Lancaster, England. Boston: MTP Press, 1988: 325.

38 Knowler WC, Barrett-Connor E, Fowler SE *et al.* Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002; 346(6): 393–403.

39 Pyne JM, Patterson TL, Kaplan RM, Gillin JC, Koch WL, Grant I. Assessment of the quality of life of patients with major depression. *Psychiatr Serv* 1997; 48(2): 224–230.

40 Andresen EM, Fouts BS, Romeis JC, Brownson CA. Performance of health-related quality-of-life instruments in a spinal cord injured population. *Arch Phys Med Rehabil* 1999; 80(8): 877–884.

41 Ganiats TG, Palinkas LA, Kaplan RM. Comparison of Quality of Well-Being scale and Functional Status Index in patients with atrial fibrillation. *Med Care* 1992; 30(10): 958–964.

42 Kaplan RM, Anderson JP, Patterson TL *et al.* Validity of the Quality of Well-Being Scale for persons with human immunodeficiency virus infection. HNRC Group. HIV Neurobehavioral Research Center. *Psychosom Med* 1995; 57(2): 138–147.

43 Bayliss MS, Gandek B, Bungay KM, Sugano D, Hsu MA, Ware Jr JE. A questionnaire to assess the generic and disease-specific health outcomes of patients with chronic hepatitis C. *Qual Life Res* 1998; 7(1): 39–55.

44 Neary MP, Cort S, Bayliss MS, Ware Jr JE. Sustained virologic response is associated with improved -related quality of life in relapsed chronic hepatitis C patients. *Semin Liver Dis* 1999; 19(Suppl. 1): 77–85.

45 Ware Jr JE, Bayliss MS, Mannocchia M, Davis GL. Health-related quality of life in chronic hepatitis C: impact of disease and treatment response. The Interventional Therapy Group. *Hepatology* 1999; 30(2): 550–555.

46 Perrillo R, Rothstein KD, Rubin R *et al.* Comparison of quality of life, work productivity and medical resource utilization of peginterferon alpha 2a vs the combination of interferon alpha 2b plus ribavirin as initial treatment in patients with chronic hepatitis C. *J Viral Hepat* 2004; 11(2): 157–165.

47 Lorig K. Outcome Measures for Health Education and Other Health Care Interventions. Thousand Oaks, CA: Sage Publications, 1996.

48 Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am J Prev Med* 1994; 10(2): 77–84.

49 Babor TF, de la Fuente JR, Saunders J, Grant M. The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Health Care. Geneva, Switzerland: World Health Organization, 1992.

50 Allen JP, Litten RZ, Fertig JB, Babor T. A review of research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcohol Clin Exp Res* 1997; 21(4): 613–619.

51 Allen JP, Reinert DF, Volk RJ. The alcohol use disorders identification test: an aid to recognition of alcohol problems in primary care patients. *Prev Med* 2001; 33(5): 428–433.

52 Reinert DF, Allen JP. The Alcohol Use Disorders Identification Test (AUDIT): a review of recent research. *Alcohol Clin Exp Res* 2002; 26(2): 272–279.

53 Gifford AL, Sengupta S. Self-management health education for chronic HIV infection. *AIDS Care* 1999; 11(1): 117–132.

54 Lorig KR, Sobel DS, Stewart AL *et al.* Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. *Med Care* 1999; 37(1): 5–14.

55 McHutchison JG, Everson GT, Gordon SC *et al.* Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *N Engl J Med* 2009; 360(18): 1827–1838.

56 Bandura A. Health promotion by social cognitive means. *Health Educ Behav* 2004; 31(2): 143–164.

57 Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. *Ann Behav Med* 2003; 26(1): 1–7.

58 Glacken M, Kernohan G, Coates V. Diagnosed with Hepatitis C: a descriptive exploratory study. *Int J Nurs Stud* 2001; 38(1): 107–116.

59 Merican I, Sherlock S, McIntyre N, Dusheiko GM. Clinical, biochemical and histological features in 102 patients with chronic hepatitis C virus infection. *Q J Med* 1993; 86(2): 119–125.

60 Kaplan RM. The minimally clinically important difference in generic utility-based measures. *COPD* 2005; 2: 91–97.

61 Coleman S, Briffa K, Conroy H, Prince R, Carroll G, McQuade J. Short and medium-term effects of an education self-management program for individuals with osteoarthritis of the knee, designed and delivered by health professionals: a quality assurance study. *BMC Musculoskelet Disord* 2008; 9: 117.

62 Lorig KR, Ritter P, Stewart AL *et al.* Chronic disease self-management program: 2-year health status and health care utilization outcomes. *Med Care* 2001; 39(11): 1217–1223.

63 Bastiaenen CH, de Bie RA, Vlaeyen JW *et al.* Long-term effectiveness and costs of a brief self-management intervention in women with pregnancy-related low back pain after delivery. *BMC Pregnancy Childbirth* 2008; 8: 19.

64 Buszewicz M, Rait G, Griffin M *et al.* Self management of arthritis in primary care: randomised controlled trial. *BMJ* 2006; 333(7574): 879.

65 Haas M, Groupp E, Muench J *et al.* Chronic disease self-management program for low back pain in the elderly. *J Manipulative Physiol Ther* 2005; 28(4): 228–237.

66 Norris SL, Engelgau MM, Narayan KM. Effectiveness of self-management training in type 2 diabetes: a systematic review of randomized controlled trials. *Diabetes Care* 2001; 24(3): 561–587.

67 Gifford AL, Laurent DD, Gonzales VM, Chesney MA, Lorig KR. Pilot randomized trial of education to improve self-management skills of men with symptomatic HIV/AIDS. *J Acquir Immune Defic Syndr Hum Retrovirol* 1998; 18(2): 136–144.

68 Patel A, Buszewicz M, Beecham J *et al.* Economic evaluation of arthritis self management in primary care. *BMJ* 2009; 339: b3532.

69 Herman WH, Hoerger TJ, Brandle M *et al.* The cost-effectiveness of lifestyle modification or metformin in preventing type 2 diabetes in adults with impaired glucose tolerance. *Ann Intern Med* 2005; 142(5): 323–332.

70 Kaplan RM, Ries AL, Reilly J, Mohsenifar Z. Measurement of health-related quality of life in the national emphysema treatment trial. *Chest* 2004; 126(3): 781–789.

71 Johnson A, Berg G, Fleegler E, Sauerbrun M. A matched-cohort study of selected clinical and utilization outcomes for an asthma care support program. *Dis Manag* 2005; 8(3): 144–154.

Potential Underuse, Overuse, and Inappropriate Use of Antidepressants in Older Veteran Nursing Home Residents

Joseph T. Hanlon, PharmD, MS,[abcdef] Xiaoqiang Wang, MS,[b] Nicholas G. Castle, PhD,[g] Roslyn A. Stone, PhD,[bh] Steven M. Handler, MD, PhD,[abcd] Todd P. Semla, PharmD, MS,[ijk] Mary Jo Pugh, PhD,[l] Dan R. Berlowitz, MD, MPH,[m] and Maurice W. Dysken, MD[n]

OBJECTIVES: To examine prevalence and resident- and site-level factors associated with potential underuse, overuse, and inappropriate use of antidepressants in older Veterans Affairs (VA) Community Living Center (CLC) residents.

DESIGN: Longitudinal study.

SETTING: One hundred thirty-three VA CLCs.

PARTICIPANTS: Three thousand six hundred ninety-two veterans aged 65 and older admitted between January 1, 2004, and June 3, 2005, with long stays (≥90 days).

MEASUREMENTS: Prevalence of potential underuse, inappropriate use, and overuse of antidepressants in residents with and without depression (as documented according to *International Classification of Diseases, Ninth Revision*, Clinical Modification, codes or Depression Rating Scale).

RESULTS: Selective serotonin reuptake inhibitors were the most commonly prescribed antidepressant. Of the 877 residents with depression, 25.4% did not receive an antidepressant, suggesting potential underuse. Of residents with depression who received antidepressants, 57.5% had potential inappropriate use due primarily to problems seen with drug–drug and drug–disease interactions. Of the 2,815 residents who did not have depression, 1,190 (42.3%) were prescribed one or more antidepressants; only 48 (4.0%) of these had a Food and Drug Administration–approved labeled indication, suggesting potential overuse. Overall, only 17.6% of antidepressant use was appropriate (324/1,844). The only consistent resident factor associated with potential underuse and overuse use was taking an antipsychotic without evidence of schizophrenia (underuse: adjusted relative risk ratio (ARRR) = 0.56, 95% confidence interval (CI) = 0.33–0.94; overuse: adjusted odds ratio = 1.52, 95% CI = 1.21–1.91). Having moderate to severe pain (ARRR = 1.54, 95% CI = 1.08–2.20) and the prescribing of an anxiolytic or hypnotic (ARRR = 1.33, 95% CI = 1.02–1.74) increased the risk of potential inappropriate antidepressant use.

CONCLUSION: Potential problems with the use of antidepressants were frequently observed in older U.S. veteran CLC residents. Future studies are needed to examine the true risks and benefits of antidepressant use in CLC and non-VA nursing homes. **J Am Geriatr Soc 59:1412–1420, 2011.**

Key words: aged; nursing homes; depression; pharmacoepidemiology

Depression is common in older nursing home residents.[1] One seminal study reported a 12% prevalence rate for major depression using the American Psychiatric Association *Diagnostic and Statistical Manual of Psychiatric Disorders, Third Edition, Revised* (DSM-III-R) criteria in older residents in a 1,100-bed nursing home.[2,3] Minor, subsyndromal, or subthreshold depression was seen in an additional 30% of these older nursing home residents.[3] In contrast, a more-recent national study of nursing homes found that only 20% of older residents had a diagnosis of

From the [a]Geriatric Research, Education and Clinical Center, and [b]Center for Health Equity Research and Promotion, Veterans Affairs Pittsburgh Health System, Pittsburgh, Pennsylvania; [c]Division of Geriatric Medicine, Department of Medicine and [d]Department of Biomedical Informatics, School of Medicine; [e]Department of Pharmacy and Therapeutics, School of Pharmacy; Departments of [f]Epidemiology, [g]Health Policy and Management, and [h]Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania; [i]Pharmacy Benefits Management Services, Department of Veterans Affairs, Hines, Illinois; Departments of [j]Medicine and [k]Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; [l]Veterans Evidence-Based Research Dissemination and Implementation Center, South Texas Veterans Affairs Health System, San Antonio, Texas; [m]Center for Health Quality, Outcomes and Economic Research, Department of Veterans Affairs, Bedford, Massachusetts; and [n]Geriatric Research, Education and Clinical Center, Department of Veterans Affairs, Minneapolis, Minnesota.

Presentations: International Conference on Pharmacoepidemiology and Therapeutic Risk Management, August 17, 2009, Providence, Rhode Island, and Gerontological Society of America Annual Meeting, November 19, 2009, Atlanta, Georgia.

Address correspondence to Dr. Joseph T. Hanlon, Department of Medicine (Geriatrics), University of Pittsburgh, Kaufman Medical Building-Suite 514, 3471 5th Ave, Pittsburgh, PA 15213. E-mail: jth14@pitt.edu

DOI: 10.1111/j.1532-5415.2011.03522.x

0002-8614/11/$15.00

depression indicated in their quarterly Minimum Data Set (MDS) assessment.[4] Depression is important to treat in older nursing home residents and is commonly associated with morbidity (e.g., hospitalization, functional status decline) and mortality.[1]

Depression in nursing homes can be treated with one or a combination of the following modes of treatment: electroconvulsive therapy, psychological or psychiatric intervention, and antidepressant therapy.[1] Antidepressant therapy is the most common treatment in nursing home residents.[1] Moreover, the prevalence of antidepressant use in U.S. nursing home residents has more than doubled—from 21.9% in 1996 to 47.5% in 2006.[5] This prevalence rate of 47.5% is consistent with the national rate of antidepressant use in Veterans Affairs (VA) Community Living Centers (CLCs).[6] Despite these high rates, data are conflicting regarding possible undertreatment of depression in nursing home residents. Recent national information shows that fewer than 5% of nursing home residents with symptoms of depression determined through the quarterly MDS assessments were not treated with an antidepressant.[7] In contrast, a 2000 study of Ohio nursing home residents found that 23% of those with a depression diagnosis did not receive an antidepressant.[8] Concomitantly, there is limited information that suggests that potential overuse and inappropriate use of antidepressants may be problematic in older nursing home residents.[9,10] Given this background, the objectives of this study were to estimate the prevalence and resident- and site-level factors associated with potential underuse, inappropriate use, and overuse of antidepressants in older VA CLC residents.

METHODS

Study Design, Setting, Data Sources, and Sample

This was a longitudinal study of 3,692 long-stay (≥90 days) residents aged 65 and older admitted to any one of the 133 VA CLCs located in the United States between January 1, 2004, and June 30, 2005. The mission of these CLCs (previously called Nursing Home Care Units) is to provide compassionate care to eligible veterans with sufficient functional impairment to require this level of service. Veterans with chronic stable conditions, including dementia, those requiring rehabilitation or short-term specialized services such as respite or intravenous therapy, and those who need inpatient hospice, can receive this type of care in VA CLCs. These CLCs are located in 21 regions across the United States called Veterans Integrated Services Networks (VISNs). The development of a merged database that included Minimum Data Set (MDS) and medication dispensing information from the Pharmacy Benefits Management Services (PBM) used for this study was recently described.[6] Briefly, CLC staff evaluated all veterans receiving care in a VA CLC using the MDS version 2.0. MDS 2.0 is a reliable standardized tool to identify the functional, psychological, and health status needs of residents and to evaluate the quality of care that these residents are receiving.[11] All MDS data were collected through resident interviews, staff interviews, and reviews of medical records. For all CLC residents, the MDS was completed at admission (within 14 days of admission), quarterly thereafter (within 90 days of previous evaluation), and at the time of any significant change in status (e.g., major change in cognitive function or functional status decline). The VA PBM provided all prescription data for the defined study cohort. These data included the start date, drug name, drug strength, dosage form, directions for use, VA therapeutic class, and amount of each drug dispensed. *International Classification of Diseases, Ninth Revision*, Clinical Modification (ICD-9-CM) codes for inpatient and outpatient diagnoses in the previous year from the VA National Patient Care Database (NPCD) records were also linked to the merged database mentioned previously. This final merged database, which was prepared using encrypted identifiers that were consistent across the three individual databases, was used to conduct the present analyses.

The sample was first stratified according to depression status determined according to ICD-9 codes. Specifically, any hospitalization or outpatient visit to a VA in the previous year during with depression was addressed was identified and noted using ICD-9-CM codes (296.2, 296.3, 298.0x, 300.4x, 309.1x, 311.xx, 301.12, 309.0x).[12,13] This approach was chosen because it was used in two previous VA studies examining the quality of depression care in outpatients and because a previous study using ICD-9 codes to identify depression found acceptable specificity (88%) but lower sensitivity (52%).[12–14] Thus, although this approach may underestimate the "true rate" of depression, it is likely to be more accurate than using just the listing of depression on a resident's problem list or in Section I of the MDS, entitled "Disease Diagnoses."[11] To ensure that those who did not have VA health service utilization in the previous year were not misclassified and to improve sensitivity, those with a high likelihood of depression on admission (MDS Depression Rating Scale (DRS) score >3)[15] were also included. The DRS is a summary of seven symptoms detected by nursing home staff that capture verbal and nonverbal indicators of depressed mood and has been shown to be a reliable (sensitivity, 91% and specificity, 69% with a psychiatrist diagnosis) and valid measure of depression in nursing home residents.[15] Eight hundred seventy-seven residents were included in the depression sample (796 according to ICD-9 codes and 181 according to DRS >3 only); the remaining 2,815 had no documented depression. The Pittsburgh VA institutional review board and research and development committees approved this study.

Main Outcome Measures

Antidepressants on the VA national formulary in 2004/05 (VA Classes CN601, CN609, CN802) included those in the following four discrete groups: tricyclic antidepressants (TCAs; amitriptyline, desipramine, doxepin, nortriptyline), selective serotonin reuptake inhibitors (SSRIs; paroxetine, sertraline, fluoxetine, citalopram), serotonin-norepinephrine reuptake inhibitors (SNRIs; venlafaxine), and other antidepressants (trazodone, mirtazapine, methylphenidate, bupropion). Methylphenidate was included because it is frequently used to treat depression in older adults.

To operationally define potential underuse and inappropriate use of antidepressants in the depression group, two specific authoritative sources were consulted: a guideline from the American Medical Directors Association (AMDA) and quality-of-care indicators from the Centers

for Medicare and Medicaid Services (CMS) for appropriate use of antidepressants for treating depression in nursing homes.[16,17] The Veterans Health Administration and Department of Defense (VHA/DOD) guideline for treating adults with depression and another from England that focused on treating older adults with depression in the primary care setting were also used.[18,19] Using a previously published and validated approach, explicit criteria for potentially inappropriate use were created that an expert panel consisting of a nurse pharmacoepidemiology researcher (MJP), a geriatric clinical pharmacist (TPS), two geriatricians (SMH, DRB), and a geriatric psychiatrist–psychopharmacologist (MWD) reviewed, edited, and agreed upon.[20] Potential inappropriate use in those in the depression group was ascertained by applying these explicit criteria to determine whether there were one or more problems in five specific quality areas: selection (e.g., choosing an antidepressant such as amitriptyline that has anticholinergic or orthostatic effects), maintenance dosage exceeding or below minimum effective dosage (e.g., highest daily dose during the 90-day period to account for the time needed to "start low and go slow" or titrate new antidepressants), clinically important drug–drug interactions; clinically important drug–disease interactions, and therapeutic duplication (use of $\geq$2 TCAs, SSRIs, or SNRIs concomitantly) (Appendix I). The lack of an order for an antidepressant during the 90-day follow-up period indicated potential underuse in the group with depression. The rationale for this operational definition is that many experts recommend antidepressant treatment for a period of time ranging from 1 to 3 years to reduce the likelihood of major depression reoccurrence and relapse in older residents with depression.[17,18] All persons in the group with depression taking an antidepressant that was not considered potentially inappropriate were included in the appropriate use category.

To operationally define potential overuse in those without depression, two specific authoritative sources were consulted: a joint statement of the members of the Long Term Care Professional Leadership Council (LTCPLC) and the Food and Drug Administration (FDA) Web site.[21,22] Potential overuse of antidepressant use in residents without depression was operationally defined as lack of a FDA-approved labeled indication. (See footnote of Table 2 for further details.)[21,22] ICD-9 codes were used to determine these indications using previously established methods.[23] Appropriate use of antidepressants in participants without depression was defined as any use not deemed to be overuse.

Independent Variables

Based on previous literature, the independent variables included demographic characteristics, health status factors, and psychiatric or neurological problems.[10,23,24] Using data from the admission MDS, categorical variables were created for age (65–74, 75–84, $\geq$85), race (black, white, or other), sex (male or female), and educational level ($<$high school, high school, $>$high school).

Regarding health status factors, a continuous variable for activity of daily living (ADL) dependencies was created from the admission MDS that had a range from 0 to 20 points and identified the amount of assistance needed from staff for five activities (bathing, dressing, grooming, toileting, and eating).[25] A continuous variable was created for the Charlson Comorbidity Index based on the methods of Deyo, which creates a score (range 0–34) calculated based on the presence of 18 chronic conditions documented in the electronic medical record using ICD-9 codes.[26,27] The number of prescribed drugs at admission was also quantified (excluding those specified below), and a dichotomous variable for physical restraint use was created as noted on the MDS. In addition, dichotomous variables for individual conditions noted on the admission MDS were examined (cancer, chronic obstructive pulmonary disease, diabetes mellitus, arteriosclerotic heart disease, arthritis, hip fracture history, hypertension, and osteoporosis).

Psychiatric and neurological problem variables were created using ICD-9 codes from VA hospitalizations or outpatient visits in the previous year. Specifically, dichotomous variables were created for cerebrovascular accident (CVA); seizure disorder; Parkinson's disease; any neuropathic pain; bipolar disease; posttraumatic stress disorder (PTSD); other anxiety disorder; and Alzheimer's, vascular, or other dementia.[27] Data from the admission MDS evaluation were also used to create a dichotomous variable for behavioral problems and moderate to severe pain and a categorical variable for cognitive function (Cognitive Performance Score (CPS): intact, mild to moderate, severe).[28,29] Finally, from PBM data, a dichotomous variable was created denoting use of individual medication classes (antipsychotics (CN701 and 709) in residents without schizophrenia, anxiolytics and hypnotics (CN302 and 309), acetylcholinesterase inhibitors (ACHEIs) and memantine (CN900)). Two dichotomous variables (bed size and geographic region) were also included to control for potentially confounding site factors.[5]

Statistical Analyses

Descriptive statistics were used to summarize independent variables and study outcomes. To include the approximately 3% of residents with missing data on education or cognitive performance status in the analyses, dummy variables were created for a "missing" category. The number and percentage of residents who were prescribed individual classes of antidepressants (TCAs, SSRIs, SNRIs, other) was described. In those with depression, the number and percentage of residents with specific types of potentially inappropriate antidepressant use were also described. A multinomial regression analysis was conducted to identify resident factors associated with underuse or inappropriate use versus appropriate use (reference group) of antidepressants in residents with depression. A backward selection approach (alpha = 0.10) was used to identify health status factors and psychiatric and neurological conditions to be added to the demographic characteristics and resident site factors in the final models. Estimated adjusted relative risk ratios (RRRs) and 95% confidence intervals (CIs) adjusted for clustering according to CLC are reported. Multiparameter Wald tests quantified the association between each outcome and the categorical variables with more than two levels. A multivariable logistic regression analysis in residents without depression was also conducted by first removing from the sample those with a FDA-approved labeled indication ("appropriate use") and the overuse

group was compared with those with no use of antidepressants.[30] Statistical analyses were performed using SAS, version 9 (SAS Institute, Inc., Cary, NC) and Stata (StataCorp, College Station, TX).

RESULTS

Table 1 compares the characteristics of CLC residents who were depressed (n = 877) with the characteristics of those who were not (n = 2,815). The groups were similar with regard to most characteristics. White residents and those with more comorbidities were more likely to be depressed than not. Those who were not depressed had more ADL dependencies and more-severe cognitive impairment than those who were depressed. The most common medication class that those without schizophrenia in both groups were taking was antipsychotics.

Table 2 summarizes antidepressant use overall and according to specific classes for residents with and without depression. The most common antidepressant class used by both groups was SSRIs. No use of monoamine oxidase inhibitors (MAOIs) was documented. Of the 877 residents with depression, 74.6% (n = 654) took an antidepressant, which suggests potential underuse in 25.4% (n = 223) of these residents. Of residents without depression, 42.3% took an antidepressant, which suggests potential overuse, because only 48 of these 1,190 taking an antidepressant had evidence of a FDA-approved labeled indication. Thus, only 4.0% of antidepressant use in those without depression was appropriate.

Table 3 summarizes potential inappropriate drug use in residents with depression. Nearly six in 10 residents with depression (n = 378) who received an antidepressant had one or more prescribing problems. Thus, appropriate antidepressant use was seen in 276 of 654 (42.5%). Drug–drug and drug–disease interactions were the most common problems, whereas therapeutic duplication and selection were the least frequent prescribing problems. By combining appropriate use regardless of depression group (48+276/1,190+654 = 17.6%), fewer than two in 10 antidepressant prescriptions were not problematic.

Table 4 summarizes the results of the multivariable multinomial logistic regression models for potential underuse and inappropriate use of antidepressant versus appropriate use in those who were depressed. Factors significantly associated with a lower risk of potential underuse in residents with depression included polypharmacy (taking >5 medications), having a history of cancer, and taking an antipsychotic without evidence of schizophrenia; the only factor associated with a greater risk of potential underuse was having ADL dependencies. Regarding potential inappropriate use, black residents and residents with cancer were significantly less likely to have this problem. Residents with moderate to severe pain and those taking an anxiolytic or hypnotic were at significantly greater risk of inappropriate use than appropriate use.

Table 5 summarizes the results of the multivariable logistic regression models for potential overuse versus no antidepressant use in residents who were not depressed. Residents aged 85 and older had a significantly lower risk of overuse, and the risk of overuse decreased with increasing comorbidity index score. Overuse was significantly more

Table 1. Patient and Facility Characteristics for Older Veterans with and without Depression in Community Living Centers

Characteristic	Depressed (n = 877)	Not Depressed (n = 2,815)
Demographic, n (%)		
Age		
65–74	265 (30.2)	869 (30.9)
75–84	468 (53.4)	1,458 (51.8)
≥85	144 (16.4)	488 (17.3)
Race		
White	760 (86.7)	2,221 (78.9)
Black	87 (9.9)	412 (14.6)
Other	30 (3.4)	182 (6.5)
Female	37 (4.2)	66 (2.3)
Education		
<High school	253 (28.8)	859 (30.5)
High school	404 (46.1)	1,312 (47.7)
>High school	208 (23.7)	614 (21.8)
Not assessed	12 (1.37)	30 (1.07)
Health status		
Number of activities of daily living dependent in, mean ± SD	8.6 (6.3)	9.3 (6.5)
Comorbidity index, mean ± SD	2.9 (2.3)	2.6 (2.2)
Number of medications other than antidepressant, n (%)		
0–5	243 (27.7)	744 (26.4)
6–10	258 (29.4)	786 (27.9)
11–15	158 (18.0)	587 (20.8)
≥16	218 (24.9)	698 (24.8)
Comorbidities, n (%)		
Chronic obstructive pulmonary disease	259 (29.5)	747 (26.5)
Diabetes mellitus	309 (35.2)	1,077 (38.3)
Cancer	162(18.5)	472 (16.8)
Arthritis	255(29.1)	756 (26.9)
Arteriosclerotic heart disease	233 (26.6)	656 (23.3)
Hip fracture	43 (4.9)	144 (5.1)
Hypertension	588 (67.0)	1,879 (66.7)
Osteoporosis	58 (6.6)	160 (5.68)
Neurological or psychiatric problems, n (%)		
Cerebrovascular accident	152 (17.3)	507 (18.0)
Seizure disorder	52 (5.9)	146 (5.19)
Parkinson's disease	80 (9.1)	146 (5.19)
Any neuropathic pain	266 (30.3)	609 (21.6)
Bipolar disease	22 (2.5)	64 (2.3)
Schizophrenia	86 (9.8)	283 (10.0)
Posttraumatic stress disorder	124 (14.1)	121 (4.3)
Other anxiety	148 (16.9)	133 (4.7)
Alzheimer's disease	106 (12.1)	286 (10.2)
Vascular dementia	89 (10.1)	150 (5.3)
Other dementia	336 (38.3)	753 (26.7)
Behavior problem	171 (19.5)	347 (12.3)
Moderate to severe pain	223 (25.4)	645 (22.9)
Cognitive function		
Intact	442 (50.4)	1,497 (53.2)
Mild to moderate impairment	325 (37.1)	880 (31.3)

(*Continued*)

Table 1. (Contd.)

Characteristic	Depressed (n = 877)	Not Depressed (n = 2,815)
Severe impairment	90 (10.3)	384 (13.6)
Not assessed	20 (2.3)	54 (1.9)
Use of antipsychotic in those without schizophrenia	214 (24.4)	512 (18.2)
Use of anxiolytic or hypnotic	68 (7.7)	163 (5.8)
Use of acetylcholinesterase inhibitor	156 (17.8)	363 (12.9)
Use of memantine	35 (4.0)	78 (2.8)
Site level indicators, n (%)		
Bed size		
Small (<60)	126 (14.4)	384 (13.6)
Medium (60–119)	416 (47.4)	1,328 (47.2)
Large (≥120)	335 (38.2)	488 (39.2)
Region		
Northeast	221 (25.2)	695 (24.7)
Midwest	191 (21.8)	556 (19.7)
South	331 (37.7)	1,158 (41.1)
West	134 (15.3)	406 (14.4)

SD = standard deviation.

likely in those with mild to moderate cognitive impairment, polypharmacy (taking >5 medications), CVA, other anxiety, and taking an antipsychotic without evidence of schizophrenia.

Table 2. Antidepressant Medication Use in Veteran Community Living Center Residents with and without Depression

Variable	n (%)	
	Depressed (n = 877)	Not Depressed (n = 2,815)
Any antidepressant use	654 (74.6)*	1,190 (42.3)†
Antidepressant class use*		
Selective serotonin reuptake inhibitor	494 (56.4)	754 (26.8)
Serotonin–norepinephrine reuptake inhibitor	44 (5.0)	42 (1.5)
Tricyclic antidepressant	32 (3.7)	87 (3.1)
Other	290 (33.1)	546 (19.4)

* Use of specific classes sums to greater than 74.6% because some patients took more than one agent concomitantly.

† Only 48 of 1,190 (4.0%) residents receiving an antidepressants had a Food and Drug Administration–approved labeled indication (venlafaxine for panic disorder, generalized anxiety disorder, and social phobia; doxepin for moderate pruritus due to atopic dermatitis or lichen simplex chronicus; bupropion for smoking cessation; methylphenidate for narcolepsy or attention deficit disorder; escitalopram for generalized anxiety disorder; fluvoxamine for social phobia or obsessive compulsive disorder; fluoxetine for obsessive compulsive disorder or panic disorder; duloxetine for diabetic peripheral neuropathy; paroxetine for generalized anxiety disorder, obsessive-compulsive disorder, panic disorder, posttraumatic stress disorder, or social phobia; sertraline for obsessive compulsive disorder, panic disorder, or posttraumatic stress disorder.

Table 3. Potentially Inappropriate Antidepressant Use in Residents with Depression According to Type of Problem and Overall (n = 877)

Type of Problem*	n (%)	Most Common Drugs Involved (n)
Selection	32 (3.7)	Amitriptyline (12)
		Nortriptyline (11)
		Doxepin (7)
Dosage	77 (8.8)	Trazodone (28)
		Sertraline (16)
		Venlafaxine (10)
Drug–drug interaction	227 (25.9)	SSRI and trazodone (73)
		Fluoxetine or paroxetine and metoprolol (41)
		Mirtazapine and SSRI (15)
Drug–disease interaction	223 (25.4)	SSRI and falls (73)
		Venlafaxine and hypertension (22)
		Tricyclic antidepressant and constipation (8)
Therapeutic	10 (1.1)	SSRI and SSRI (10)
Duplication		
Any problem	378 (43.1)	

* Sums to more than 43.1% because some residents had more than one type of problem.

SSRI = selective serotonin reuptake inhibitor.

DISCUSSION

In this study, nearly 50% of all older long-stay veteran nursing home residents received an antidepressant, which is consistent with the rate of nearly 48% of non-VA nursing home residents taking an antidepressant.[6] That depression was found in nearly 25% of residents is also consistent with previously published studies,[3,10] although it was found that nearly 25% of those with depression did not receive an antidepressant, suggesting potential underuse. This rate is considerably less than the 45% of nursing home residents with MDS-reported depression who were not being given an antidepressant in a multistate U.S. sample,[10] but it is consistent with the rates from more-recent studies that show that between 21% and 34% of nursing home residents with depression do not receive an antidepressant.[8,31] The multivariable analyses of factors associated with underuse of antidepressants suggest that prescribers may be more cautious in residents with greater ADL dependencies. This may reflect appropriate concern that the likelihood of antidepressant adverse effects is greater than the potential benefits in these vulnerable people. It is hoped that better detection and monitoring of depression using the valid, reliable, and frequently used nine-item Patient Health Questionnaire, which is replacing the DRS in MDS version 3.0 and is scheduled to be implemented in non-VA nursing homes in the fall of 2010 and VA CLCs in 2011, will further reduce the rate of antidepressant underuse.[32]

Of persons who were depressed and receiving an antidepressant, nearly 60% had evidence of potentially inappropriate use, with one or more prescribing problems. The least-frequent problems were therapeutic duplication and selection. Medication selection was potentially

Table 4. Comparison of Factors Associated with Underuse (n = 223) and Inappropriate Use (n = 378) and Those Associated with Appropriate Use (Reference Group; n = 276) of Antidepressants in Residents with Depression

Factor	Adjusted Relative Risk Ratio (95% Confidence Interval)	
	Underuse (n = 223)	Inappropriate Use (n = 378)
Demographic		
Age (reference 65–74)		
75–84	0.90 (0.59–1.39)	1.27 (0.87–1.84)
≥85	0.92 (0.52–1.62)	1.29 (0.79–2.11)
Race (reference white)		
Black	0.85 (0.49–1.49)	0.48 (0.30–0.76)*
Other	1.21 (0.48–3.01)	0.86 (0.33–2.25)
Female gender	1.55 (0.68–3.53)	0.67 (0.28–1.61)
Education (reference <high school)		
High school	0.84 (0.55–1.28)	0.89 (0.59–1.33)
>High school	1.02 (0.58–1.79)	0.90 (0.57–1.44)
Not assessed	1.15 (0.27–4.81)	1.23 (0.26–5.83)
Health status		
Activity of daily living score (per unit increase)	1.05 (1.02–1.09)*	1.02 (0.99–1.04)
Number of medications other than antidepressant (reference 0–5)		
6–10	0.57 (0.36–0.91)*	1.39 (0.88–2.19)
11–15	0.40 (0.23–0.73)*	1.58 (0.94–2.66)
≥16	0.46 (0.28–0.76)*	1.79 (1.09–2.94)
Cancer	0.52 (0.33–0.81)*	0.62 (0.41–0.94)*
Neurological or psychiatric problem		
Cerebrovascular accident	0.63 (0.37–1.08)	1.33 (0.83–2.15)
Behavior problem	1.51 (0.91–2.49)	0.69 (0.44–1.08)
Moderate to severe pain	0.79 (0.51–1.21)	1.54 (1.08–2.20)*
Use of anxiolytic or hypnotic	1.08 (0.81–1.44)	1.33 (1.02–1.74)*
Use of antipsychotic in resident without schizophrenia	0.56 (0.33–0.94)*	0.90 (0.62–1.30)
Site-level indicators		
Bed size (reference small (<60))		
Medium (60–120)	0.90 (0.49–1.66)	0.97 (0.56–1.66)
Large (>120)	0.59 (0.32–1.11)	1.01 (0.60–1.71)
Region (reference Northeast)		
Midwest	0.76 (0.44–1.29)	0.78 (0.47–1.31)
South	0.61 (0.38–0.99)	1.03 (0.66–1.60)
West	0.77 (0.43–1.37)	1.28 (0.74–2.19)

* $P < 0.05$; for categorical variables, contrasts are noted as being statistically significant only when the overall effect in the equation is significant.
Wald chi-square (46) = 147.57; probability > chi-square = 0.0000; log pseudolikelihood = −871.022; pseudo coefficient of determination = 0.076.

Table 5. Comparison of Factors Associated with Overuse (n = 1,142) and No Use (Reference Group; n = 1,625) in Residents without Depression*

Factor	Adjusted Odds Ratio (95% Confidence Interval)
Demographic	
Age (reference 65–74)	
75–84	0.89 (0.73–1.09)
85+	0.70 (0.57–0.87)†
Race (reference white)	
Black	0.82 (0.65–1.03)
Other	0.69 (0.40–1.19)
Female	1.28 (0.81–2.01)
Education (reference <high school)	
High school	1.08 (0.88–1.33)
>High school	1.32 (1.05–1.68)
Not assessed	0.63 (0.27–1.43)
Health status	
Comorbidity index	0.92 (0.88–0.96)†
Number of medications other than antidepressant (reference 0–5)	
6–10	1.88 (1.48–2.38)†
11–15	2.50 (1.93–3.24)†
≥16	3.50 (2.79–4.38)†
Cancer	1.27 (0.99–1.63)
Chronic obstructive pulmonary disease	1.21 (1.00–1.47)
Arteriosclerotic heart disease	1.20 (0.96–1.50)
Neurological or psychiatric problem	
Cerebrovascular accident	1.50 (1.20–1.87)†
Any neuropathic pain	1.17 (0.98–1.40)
Posttraumatic stress disorder	1.09 (0.67–1.77)
Other anxiety	1.48 (1.02–2.14)†
Cognitive function	
Intact	1.00 (Reference
Mild to moderate impairment	1.24 (1.02–1.50)†
Severe impairment	0.96 (0.72–1.27)
Not assessed	1.75 (0.97–3.16)
Use of antipsychotic in residents without schizophrenia	1.52 (1.21–1.91)†
Site-level indicators	
Bed size (reference small (<60))	
Medium (60–120)	0.84 (0.60–1.18)
Large (>120)	1.02 (0.72–1.45)
Region (reference Northeast)	
Midwest	1.13 (0.77–1.65)
South	1.16 (0.80–1.69)
West	1.03 (0.69–1.53)

* Those with appropriate on-label antidepressant use (n = 48) excluded from the model.
† $P < .05$.
Wald chi-square (28) = 251.75; probability > chi-square = 0.000; log pseudolikelihood = −1761.431; pseudo coefficient of determination = 0.0612.

problematic primarily because TCAs are notorious for causing orthostatic hypotension and having both of anticholinergic effects, which can increase the risk of falls and cognitive impairment in older adults.[16–18] Under- and overdosing problems were seen in nearly 9% of residents with depression. Underdosing was most common with sertraline, trazodone, and venlafaxine. Trazodone may have been misclassified as underdosed because it may have been prescribed to manage sleep and weight loss, despite little evidence-based data to support these indications.[33] Drug–drug interactions were seen in one in four antidepressant users who were depressed. The most common drug–drug interactions were the use of multiple drugs that increase

serotonin (and thus increase the risk of serotonin syndrome); this would include the use of multiple antidepressants regardless of therapeutic intent.[34] The next most common drug–drug interactions involved the use of paroxetine, fluoxetine, or bupropion, which are potent inhibitors of CY2D6 hepatic enzymes, in combination with important substrate drugs such as metoprolol and other antidepressants (TCAs, venlafaxine), which could result in preventable adverse drug events.[35] Drug–disease interactions were just as common in this resident group and frequently involved the prescribing of antidepressants in residents with a history of a fall. The risk of falls with SSRIs is the same as that with TCAs.[36] The only potentially modifiable risk factors associated with potential inappropriate prescribing of antidepressants in this study were residents with moderate to severe pain and the prescribing of an anxiolytic or hypnotic.

To the best of the knowledge of the authors, this is one of the first studies to examine potential overuse of antidepressants in nursing home residents. In residents without depression, only a small number (48/1,190) had a FDA-approved labeled indication for the antidepressants. One explanation is that a recent study showed that U.S. physicians have limited knowledge of which indications are FDA approved versus being off-label.[37] Of potential concern is the recent report that five antidepressants are among the top 25 drugs used off label with inadequate efficacy evidence.[38] One factor associated with potential overuse was anxiety, for which there is evidence that specific antidepressants classes (and not just individual agents) may be effective; this use is supported by various nursing home organizations.[21] Finally, coprescribing of antipsychotics (in residents without schizophrenia) was associated with greater risk of antidepressant overuse.

So what are the implications of these results? One is that that there are prescribing quality problems involving antidepressants that clinicians should be aware of in VA CLCs. It is likely that similar prescribing problems are also occurring at similar levels in non-VA nursing homes given their equally high rates of antidepressant use.[5] What is not clear is the effect that this antidepressant prescribing quality has on nursing home resident outcomes. Nonetheless, it is clinically sensible to consider ways to address this quality prescribing problem. Three recently published articles describe successful approaches used in randomized controlled trials (academic detailing, pharmacist interventions, multidisciplinary teamwork, computerized decision support systems) to improve prescribing of psychotropic medications for nursing home residents,[39–41] although none of these studies examined changing the quality of antidepressant prescribing. In part to address this concern, the VA is launching a variety of initiatives, including increasing the availability and integration of psychology and psychiatric services in CLCs and increasing staff education. Similar initiatives in non-VA nursing homes will be also be necessary to address the stigma associated with diagnosing and treating psychiatric problems in nursing homes and historically low reimbursement rates for nonpsychiatrist providers.

This study has a number of potential limitations. There is potential misclassification because an independent research psychiatrist did not diagnose depression. Instead, those with depression were classified according to ICD-9 codes or severe depressive symptoms based on MDS data. Examining alternative classifications of depression, including shortening the lookback period for ICD-9 codes to 6 months and using the listing of depression in the MDS, did not substantially change the depression sample. The application of explicit criteria to evaluate the quality of prescribing is limited because they cannot take into account individual resident characteristics. In addition, the rate of potential underuse may be somewhat inflated because residents may have been receiving effective nonpharmacological treatment that this and other studies did not capture. Some explicit guideline criteria published in 2006 or later were also applied to data from 2004/05, which does not allow for prior dissemination of this information to providers. Finally, it is unclear what the generalizability of the current findings are to non-VA nursing home settings, given that the majority of their residents are older women and that the use of some antidepressant medications may be different in VA because of their use of a national formulary.

Despite these potential limitations, potential problems with the use of antidepressants were observed frequently in older U.S. veteran nursing home residents. Future studies are needed to examine the true risks and benefits of antidepressant use in nursing homes.

ACKNOWLEDGMENTS

Conflict of Interest: This study was supported by National Institute of Aging Grants P30AG024827, T32 AG021885, K07AG033174, R01AG034056, R56AG027017, and 3U01 AG012553; National Institute of Mental Health Grant R34 MH082682; National Institute of Nursing Research Grant R01 NR010135; Agency for Healthcare Research and Quality Grants R01 HS017695, R01 HS018721, and K12 HS019461; and VA Health Services Research Grant IIR-06-062.

Author Contributions: Dr. Hanlon conceived of and designed the study, acquired the data, supervised the analyses and interpretation of the data, and drafted the initial manuscript. Mr. Wang performed the analyses and assisted in the interpretation of data and preparation of the manuscript. Dr. Castle served as an expert on the creation and use of MDS scales as important covariates; contributed to the design, analyses, and interpretation of data for this study; and assisted in preparing the manuscript. Drs. Handler, Semla, Pugh, Berlowitz, and Dysken served on the expert panel that created the explicit criteria for evaluating potentially inappropriate antidepressant use, and all contributed to the design, analyses, and interpretation of data for this study and assisted in preparing the manuscript. Dr. Stone assisted in the development of the study design and analytical plan, oversaw the statistical analyses performed by Mr. Wang, contributed to the interpretation of data for this study, and assisted in preparing the manuscript.

Sponsor's Role: The organizations that funded this study had no role in the design, methods, analysis, or preparation of this paper.

REFERENCES

1. Thakur M, Blazer DG. Depression in long term care. J Am Med Dir Assoc 2008;9:82–87.
2. Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised. Washington, DC: American Psychiatric Association, 1987.

3. Parmelee PA, Katz IR, Lawton MP. Depression among institutionalized aged: Assessment and prevalence estimation. J Gerontol 1989;44:M22–M29.
4. Jones RN, Marcantonio ER, Rabinowitz T. Prevalence and correlates of recognized depression in U.S. nursing homes. J Am Geriatr Soc 2003;51: 1404–1409.
5. Hanlon JT, Handler SM, Castle NG. Antidepressant prescribing in U.S. nursing homes between 1996–2006 and its relationship to staffing patterns and use of other psychotropic medications. J Am Med Dir Assoc 2010;11:320–324.
6. Hanlon JT, Wang X, Good CB et al. Racial differences in medication use among older long stay veteran nursing home care unit patients. Consult Pharm 2009;24:439–446. PMCID: PMC2734488.
7. Centers for Medicare and Medicaid Services (CMS). The MDS Quality Indicator (QI). [on-line]. Available at http://www.cms.gov/MDSPubQIand ResRep/03_qireports.asp, 2005 Accessed July 15, 2010.
8. Levin CA, Wei W, Akincigil A et al. Prevalence and treatment of diagnosed depression among elderly nursing home resident in Ohio. J Am Med Dir Assoc 2007;8:585–594.
9. Heston LL, Garrard J, Makris L et al. Inadequate treatment of depressed nursing home elderly. J Am Geriatr Soc 1992;40:1117–1122.
10. Brown MN, Lapane KL, Luisi AF. The management of depression in older nursing home residents. J Am Geriatric Soc 2002;50:69–76.
11. Mor V. A comprehensive clinical assessment tool to inform policy and practice: Applications of the Minimum Data Set. Med Care 2004;42(4 Suppl):III50–III59.
12. Charbonneau A, Rosen AK, Ash AS et al. Measuring the quality of depression care in a large integrated health system. Med Care 2003;41:669–680.
13. Jordan N, Lee TA, Valenstein M et al. Effect of care setting on evidence-based depression treatment for veterans with COPD and comorbid depression. J Gen Int Med 2007;22:1447–1452.
14. Spetell CM, Wall TC, Allison J et al. Identifying physician recognized depression from administrative data. Health Serv Res 2003;38:1081–1102.
15. Burrows AB, Morris JN, Simon SE et al. Development of a Minimum Data Set-based depression rating scale for use in nursing homes. Age Ageing 2000; 29:165–172.
16. Centers for Medicare and Medicaid Services (CMS). State Operations Manual: Surveyor Guidance for Unnecessary Medications (F329). [on-line]. Available at http://www.cms.hhs.gov/transmittals/downloads/R22SOMA.pdf Accessed July 15, 2010.
17. American Medical Directors (AMDA). Guidelines regarding the use of antidepressants. The American Medical Directors, 2003.
18. Veterans Health Administration and Department of Defense (VHA/DOD) Clinical Practice Guideline: Management of Major Depressive Disorder: Veterans Health Administration and Department of Defense, 2009 [on-line]. Available at http://www.healthquality.va.gov/Major_Depressive_Disorder_ MDD_Clinical_Practice_Guideline.asp Accessed July 15, 2010.
19. Baldwin RC, Anderson D, Black S et al. Guideline for the management of late life depression in primary care. Int J Geriatr Psych 2003;18:829–838.
20. Hanlon JT, Schmader KE, Boult C et al. Use of inappropriate prescription drugs by older people. J Am Geriatr Soc 2002;50:26–34.
21. Long Term Care Professional Leadership Council, American College of Health Care Administrators, American Medical Directors Association, American Society of Consultant Pharmacists, National Association Directors of Nursing Administration/Long Term Care. Use of antidepressants in nursing home residents. A joint statement of the members of the Long Term Care Professional Leadership Council (LTCPLC). Consult Pharm 2008;23:231–234.
22. Drugs@ FDA. [on-line]. Available at http://www.accessdata.fda.gov/scripts/ cder/drugsatfda Accessed July 15, 2010.
23. Pugh MJ, Fincke BG, Bierman AS et al. Potentially inappropriate prescribing in elderly veterans: Are we using the wrong drug, wrong dose, or wrong duration? J Am Geriatr Soc 2005;53:1282–1289.
24. Wang PS, Schneeweiss S, Brookhart MA et al. Suboptimal antidepressant use in the elderly. J Clin Psychopharmacol 2005;25:118–126.
25. McConnell ES, Pieper CF, Sloane RJ et al. Effects of cognitive performance on change in physical function in long-stay nursing home residents. J Gerontol A Biol Sci Med Sci 2002;57A:M778–M784.
26. Charlson ME, Pompei P, Ales KL et al. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis 1987;40:373–383.
27. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992; 45:613–619.
28. Liperoti R, Mor V, Lapane KL et al. The use of atypical antipsychotics in nursing homes. J Clin Psychiatry 2003;64:1106–1112.
29. Hartmaier SL, Sloane PD, Guess HA et al. Validation of the Minimum Data Set Cognitive Performance Scale: Agreement with the Mini-Mental State Examination. J Gerontol A Biol Sci Med Sci 1995;50A:M128–M133.
30. Agresti A. Categorical Data Analysis, 2nd Ed. Hoboken, NJ: Wiley-Interscience, 2002.
31. Damian J, Pstor-Barriuso R, Valderramaa-Gama E. Descriptive epidemiology of undetected depression in institutionalized older people. J Am Med Dir Assoc 2010;11:312–319.
32. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 2001;16:606–613.
33. Kamel NS, Gammack JK. Insomnia in the elderly: Cause, approach, and treatment. Am J Med 2006;119:463–469.
34. Hall RCW, Hall RCW, Chapman MJ. Central serotonin syndrome. Clin Geriatr 2007;15:18–24.
35. Spina E, Santoro V, D'Arrigo C. Clinically relevant pharmacokinetic drug interactions with second generation antidepressants: An update. Clin Ther 2008;30:1206–1227.
36. Hartikainen SA, Lonnroos E, Louhivuori K. Medication as a risk factor for falls: Critical systematic review. J Gerontol A Biol Sci Med Sci 2007;62A: 1172–MS1181.
37. Chen DT, Wynia MK, Moloney RM et al. US physician knowledge of the FDA approved indications and evidence base for commonly prescribed drugs: Results of a national survey. Pharmacoepi Drug Safety 2009;18: 1094–1100.
38. Walton SM, Schumock GT, Alexander GC et al. Prioritizing future research on off label prescribing: Results of a quantitative evaluation. Pharmacother 2008;28:1443–1452.
39. Nishtala PS, McLachlan AJ, Bell JS et al. Psychotropic prescribing in long-term care facilities: Impact of medication reviews and educational interventions. Am J Geriatr Psychiatry 2008;16:621–632.
40. Marcum ZA, Handler SM, Wright R et al. Interventions to improve suboptimal prescribing in nursing homes. Am J Geriatr Pharmacother 2010;8:183–200.
41. Donovan JL, Kanaan AO, Thomson MS et al. Effect of clinical decision support on psychotropic medication prescribing in the long-term care setting. J Am Geriatr Soc 2010;58:1005–1007.

APPENDIX I

Table A1. Explicit Criteria for Antidepressant Use in Older Nursing Home Residents

Class and Agent	Selection	Minimum/Maximum Daily Dosage (mg/d)	Drug–Drug Interaction to Avoid	Drug–Disease Interactions*	Therapeutic Duplication
Miscellaneous antidepressant					
Bupropion	Recommended	150–300	CYP2D6 substrates[†]	Seizure disorder	NA
Mirtazapine	Recommended	15–45 (30 if estimated creatinine clearance <30 mL/min)	Clonidine, other drugs that ↑ serotonin[‡]	None	NA
Trazodone	Recommended	25–150	Other drugs that ↑ serotonin[‡]	None	NA
Methylphenidate	Recommended	5–20	Monoamine oxidase inhibitors	Hypertension, seizure disorder, arrhythmia, long QT interval	Other amphetamines and modafinil
Serotonin-norepinephrine reuptake inhibitor					
Venlafaxine	Recommended	50–225	Other drugs that ↑ serotonin[‡]	Hypertension	NA
SSRI					
Citalopram	Recommended	10–40	Other drugs that ↑ serotonin[‡]	Falls	Concurrent SSRI
Fluoxetine	Recommended	10–40	CYP2D6 substrates,[†] other drugs that ↑ serotonin,[‡] phenytoin	Falls	Concurrent SSRI
Paroxetine	Recommended	10–40	Anticholinergics,[‡] CYP2D6 substrates,[†] other drugs that ↑ serotonin[‡]	Falls	Concurrent SSRI
Sertraline	Recommended	50–200	Other drugs that ↑ serotonin[‡]	Falls	Concurrent SSRI
TCA					
Amitriptyline	Not recommended[1]	10–75	Anticholinergic,[§] bupropion, clonidine, other drugs that ↑ serotonin[‡]	Benign prostatic hypertrophy, constipation, dementia, falls, heart block, orthostatic hypotension	Concurrent TCA
Desipramine	Recommended	10–75	Anticholinergic,[§] bupropion, clonidine, other drugs that ↑ serotonin[‡]	Benign prostatic hypertrophy, constipation, dementia, falls, heart block, orthostatic hypotension	Concurrent TCA
Doxepin	Not recommended	10–75	Anticholinergic,[§] bupropion, clonidine, other drugs that ↑ serotonin[‡]	Benign prostatic hypertrophy, constipation, dementia, falls, heart block, orthostatic hypotension	Concurrent TCA
Nortriptyline	Recommended	10–75	Anticholinergic,[§] bupropion, clonidine, other drugs that ↑ serotonin[‡]	Benign prostatic hypertrophy, constipation, dementia, falls, heart block, orthostatic hypotension	Concurrent TCA

* Diseases were determined from admission Minimum Data Set (version 2.0) assessments and through the use of specific *International Classification of Diseases, Ninth Revision*, codes. Although this approach may not be highly sensitive, it is likely to be highly specific.

† CYP2D6 substrates (metoprolol, tricyclic antidepressants, venlafaxine).

‡ Other nonantidepressant drugs that increase serotonin that in combination with specific antidepressants increase the risk of serotonin syndrome (buspirone, dextromethorphan, meperidine, sumatriptan, tramadol).

§ Nonantidepressant drugs with anticholinergic activities included antiarrhythmic (disopyramide), anti-emetic and anti-vertigo (meclizine, prochlorperazine), antiparkinsonian (trihexyphenidyl), antipsychotic (all conventional antipsychotics, olanzapine, quetiapine), antispasmodic (e.g., belladonna, oxybutynin), cold and allergy drug (e.g., hydroxyzine and other first-generation antihistamines), sleep aid (diphenhydramine), and skeletal muscle relaxant (cyclobenzaprine and methocarbamol).

SSRI = selective serotonin reuptake inhibitor; TCA = tricyclic antidepressant.

Relation between Sex Hormone Concentrations, Peripheral Arterial Disease, and Change in Ankle-Brachial Index: Findings from the Framingham Heart Study

Robin Haring, Thomas G. Travison, Shalender Bhasin, Ramachandran S. Vasan, Henri Wallaschofski, Maithili N. Davda, Andrea Coviello,* and Joanne M. Murabito*

Preventive Medicine and Epidemiology Section (R.H., R.S.V., A.C.); Sections of Endocrinology, Diabetes, and Nutrition (T.G.T., S.B., M.N.D., A.C.) and General Internal Medicine (J.M.M.); and Department of Biostatistics (T.G.T.), Boston University School of Public Health, Boston University School of Medicine, Boston, Massachusetts 02118; Institute of Clinical Chemistry and Laboratory Medicine (R.H., H.W.), University of Greifswald, 17475 Greifswald, Germany; and National Heart, Lung, and Blood Institute's Framingham Heart Study (R.S.V., J.M.M.), Framingham, Massachusetts 01701

Objective: Our objective was to investigate cross-sectional and longitudinal associations of sex hormone concentrations with ankle-brachial index (ABI) and peripheral arterial disease (PAD).

Methods and Results: We used data from 3034 (1612 women) participants of the Framingham Heart Study. ABI was measured and PAD defined as ABI below 0.90, intermittent claudication, or lower extremity revascularization. Sex hormone concentrations were measured by liquid chromatography-tandem mass spectrometry [total testosterone (T), total estradiol, and estrone], immunofluorometric assay (SHBG), or calculated (free T). Sex-specific multivariable linear and logistic regression models were conducted for each sex hormone separately. Cross-sectional multivariable analyses revealed that men with lower free T and higher estrone (E1) concentrations had a significantly lower ABI [for free T, lowest *vs.* higher quartiles, $\beta = -0.02$, with 95% confidence interval (CI) $= -0.04$ to -0.001; and for E1, highest *vs.* lower quartiles, $\beta = -0.02$, with 95% CI $= -0.04$ to -0.002, respectively). Lower total T and SHBG concentrations were also associated with prevalent PAD in age-adjusted [odds ratio (OR) = 2.24, 95% CI = 1.17–4.32; and OR = 2.06; 95% CI = 1.07–3.96, lowest *vs.* highest quartile, respectively), but not in multivariable logistic regression models. Longitudinal multivariable analyses showed an association of lower SHBG with ABI change (decline ≥ 0.15; n = 69) in men [OR for SHBG quartiles 1, 2, and 3 as compared with quartile 4 were 2.56 (95% CI = 1.01–6.45), 2.28 (95% CI = 0.98–5.32), and 2.93 (95% CI = 1.31–6.52), respectively]. In women, none of the investigated associations yielded statistically significant estimates.

Conclusion: Our investigation of a middle-aged community-based sample suggests that sex hormone concentrations in men but not in women may be associated with PAD and ABI change. **(*J Clin Endocrinol Metab* 96: 3724–3732, 2011)**

Peripheral artery disease (PAD) is one of the most common manifestations of atherosclerosis, affecting about 27 million individuals in Europe and North America (1). PAD is a powerful and independent risk factor of cardiovascular morbidity and mortality (2–4). As an early indicator of PAD, a low ankle-brachial index (ABI) has also been associated with increased risk of subsequent cardiovascular disease (CVD) and mortality (5). Several pro-

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in U.S.A.

doi: 10.1210/jc.2011-1068 Received March 23, 2011. Accepted August 29, 2011.
First Published Online September 20, 2011

* A.C. and J.M.M. contributed equally.

Abbreviations: ABI, Ankle-brachial index; CI, confidence interval; CV, coefficient of variation; CVD, cardiovascular disease; E1, estrone; E2, estradiol; FHS, Framingham Heart Study; IC, intermittent claudication; LC-MS/MS, liquid chromatography-tandem mass spectrometry; OR, odds ratio; PAD, peripheral arterial disease; TT, total testosterone.

spective investigations have shown that low total testosterone (TT) concentrations in men were associated with a less favorable cardiovascular risk profile including obesity (6), incident metabolic syndrome (7), diabetes mellitus (8), dyslipidemia (9), hypertension (10), and mortality (11). In women, previous studies of the relation between T, SHBG, and CVD have yielded conflicting results (12). However, most studies in women suggest that higher T and lower SHBG concentrations are associated with an adverse CVD risk factor profile including visceral fat accumulation (13), insulin resistance, adverse lipid profiles (14), diabetes (15), subclinical atherosclerosis (16–18), and increased risk of incident CVD (19, 20).

Given the suggested associations of sex hormones, ABI, and PAD with cardiovascular risk factors, morbidity, and mortality, it is intriguing that data relating circulating sex hormone concentrations to ABI and PAD are very limited. To date, there is only one cross-sectional study in elderly men reporting a positive correlation between low free T concentrations and prevalent PAD (21). However, cross-sectional studies are limited in their ability to assess causality, and therefore, no directionality for the observed association can be inferred from these studies. Thus, evidence for a prospective association of sex hormones with PAD is lacking to date. Accordingly, we investigated the cross-sectional and longitudinal associations of circulating sex hormone concentrations with ABI and PAD in the community-based Framingham Heart Study (FHS) Offspring cohort.

Subjects and Methods

Study population

The FHS Offspring Study was initiated in 1971 to examine 5124 adult children (and offspring spouses) of the original FHS cohort approximately every 4–8 yr (22). Written informed consent was obtained at each examination, and the Institutional Review Board of the Boston University Medical Center approved the examination content. From the 3334 FHS participants participating in in-person clinical evaluations at the seventh Offspring examination cycle (1998–2001), we excluded individuals missing sex hormone data due to insufficient stored serum (n = 176), missing ABI at examination 7 (n = 83), or with ABI higher than 1.40 at examination 7 (n = 15); men reporting use of medications that could influence sex hormones, such as leuprolide for prostate cancer or T replacement (n = 4); and women taking oral contraceptives (n = 19) or for whom menopause status was indeterminate at examination 7 (n = 3). Due to their large proportion (31.6%), postmenopausal women on hormone therapy were not excluded, and we performed subsidiary analyses that included this subgroup. Our final study population comprised 1422 men and 1612 women; of these individuals, 2473 had ABI measured at examination 8 (2005–2008) and were eligible for longitudinal analyses of ABI change between the two examinations (Fig. 1).

FIG. 1. Study population flow chart indicating the number of subjects and exclusions separately for men and women. FT, Free T; HT, hormone therapy; OC, oral contraceptive.

Sex hormone measurements

Serum samples were drawn from the antecubital vein in the supine position between 0800 and 0900 h, after an overnight fast of about 10 h. The samples were aliquoted and immediately stored at −80 C and remained frozen until the time of assay. TT, total estradiol (E2), and estrone (E1) concentrations were measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS) as previously described (23, 24). As part of the Centers for Disease Control's Testosterone Assay Harmonization Initiative, quality control samples were run for TT after each batch of 200 study samples. In addition, 28 serum samples from men and women with TT concentrations across the entire male

TABLE 1. Baseline characteristics of the study population

Variable	Men (n = 1422)	Premenopausal women (n = 217)	Postmenopausal women (+ HT) (n = 510)	Postmenopausal women (− HT) (n = 885)
Age (yr)	61.0 ± 9.5	49.1 ± 4.8	60.2 ± 7.1	64.6 ± 8.5
Waist circumference (cm)	103.4 ± 11.2	92.8 ± 16.3	94.6 ± 14.5	98.7 ± 14.9
Body mass index (kg/m^2)	28.8 ± 4.4	27.5 ± 6.6	27.0 ± 5.5	27.8 ± 5.5
Serum TT (nmol/liter)	20.29 ± 8.02	1.19 ± 0.6	1.05 ± 0.56	1.06 ± 0.77
Serum Free T (nmol/liter)	0.30 ± 0.12	0.01 ± 0.007	0.01 ± 0.005	0.01 ± 0.009
Serum SHBG (nmol/liter)	58.55 ± 27.33	87.26 ± 52.67	140.46 ± 72.16	72.01 ± 36.55
Total estradiol (pmol/liter)	97.16 ± 33.48	414.38 ± 430.04	135.95 ± 174.81	40.78 ± 53.35
Estrone (pmol/liter)	189.74 ± 67.39	361.37 ± 297.92	628.28 ± 743.51	116.16 ± 65.82
ABI	1.14 ± 0.12	1.12 ± 0.08	1.10 ± 0.10	1.08 ± 0.11
ABI <0.90	64 (4.5%)	2 (0.9%)	10 (2.0%)	46 (5.2%)
IC	51 (3.6%)	1 (0.5%)	16 (3.1%)	25 (2.8%)
Lower extremity revascularization	9 (0.63%)			2 (0.23%)
Clinical PAD	89 (6.3%)	3 (1.4%)	19 (3.7%)	59 (6.7%)
Current smoking	161 (11.5%)	29 (13.4%)	53 (10.5%)	106 (12.2%)
Total cholesterol (mg/dl)	193.2 ± 35.1	200.6 ± 36.0	205.4 ± 34.3	210.3 ± 38.1
HDL cholesterol (mg/dl)	45.6 ± 12.9	61.7 ± 16.3	65.6 ± 17.0	58.2 ± 16.9
Statin use	316 (22.2%)	11 (5.1%)	68 (13.3%)	172 (19.4%)
Systolic blood pressure (mm Hg)	127.6 ± 16.9	115.6 ± 13.4	123.7 ± 18.1	127.7 ± 19.9
Diastolic blood Pressure (mm Hg)	75.8 ± 9.3	73.8 ± 8.8	72.3 ± 8.4	71.8 ± 9.5
Antihypertensive medication	521 (36.7%)	27 (12.4%)	166 (32.6%)	301 (34.0%)
Hypertension	556 (39.1%)	30 (13.9%)	176 (34.5%)	330 (37.4%)
Diabetes	191 (13.5%)	7 (3.2%)	29 (5.7%)	99 (11.2%)
CVD	218 (15.3%)	3 (1.4%)	34 (6.7%)	77 (8.7%)

Data are absolute number (percentages) or mean ± SD. To convert the values of TT from nanomoles per liter to nanograms per deciliter, multiply by 28.82. To convert the values for free T to picograms per milliliter, divide by 3.467. To convert estradiol and estrone from picomoles per liter to picograms per milliliter, divide the estradiol value by 3.671 and the estrone value by 3.699. Clinical PAD was defined as a composite outcome including an ABI lower than 0.90, IC, or lower extremity revascularization. HDL, High-density lipoprotein cholesterol; HT, hormone therapy.

and female range were measured in a blinded manner in the Boston University and Mayo Clinics laboratories. Pearson correlation coefficient (>0.99) and Bland-Altman plots revealed no significant differences between values obtained in the two laboratories at any concentration (25). The functional sensitivity of the assay was 2 ng/dl, and interassay coefficients of variation (CV) were 10.6% at 0.82 nmol/liter (23.5 ng/dl), 7.9%, at 1.7 nmol/liter (48.6 ng/dl), 7.7% at 8.4 nmol/liter (241 ng/dl), 4.4%

TABLE 2. Cross-sectional associations of sex hormone concentrations in men with ABI and clinical PAD

	TT		Free T		SHBG	
	ABI, β (95% CI)	PAD, OR (95% CI)	ABI, β (95% CI)	PAD, OR (95% CI)	ABI, β (95% CI)	PAD, OR (95% CI)
Age-adjusted models						
<25th	−0.02 (−0.03, 0.001)	2.24 (1.17, 4.32)[b]	−0.02 (−0.04, −0.01)[b]	1.92 (0.96, 3.87)	−0.005 (−0.02, 0.01)	2.06 (1.07, 3.96)[b]
25–50th	−0.005 (−0.02, 0.01)	1.37 (0.68, 2.77)	−0.009 (−0.03, 0.01)	1.57 (0.76, 3.23)	0.007 (−0.01, 0.02)	1.24 (0.65, 2.38)
50–75th	−0.004 (−0.02, 0.01)	1.48 (0.73, 2.98)	−0.001 (−0.02, 0.02)	0.95 (0.42, 2.12)	0.005 (−0.01, 0.02)	1.50 (0.82, 2.74)
>75th	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
P for Trend	0.25	0.08	0.04	0.11	0.58	0.17
Multivariable-adjusted models[a]						
<25th	−0.01 (−0.03, 0.01)	1.50 (0.70, 3.21)	−0.02 (−0.04, −0.001)[b]	1.29 (0.58, 2.86)	−0.001 (−0.02, 0.02)	1.69 (0.80, 3.60)
25–50th	0.0002 (−0.02, 0.02)	0.82 (0.37, 1.79)	−0.006 (−0.02, 0.01)	1.25 (0.57, 2.79)	0.007 (−0.01, 0.02)	1.02 (0.50, 2.11)
50–75th	−0.003 (−0.02, 0.01)	1.09 (0.51, 2.35)	0.001 (−0.02, 0.02)	0.71 (0.29, 1.72)	0.003 (−0.01, 0.02)	1.36 (0.70, 2.64)
>75th	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
P for trend	0.53	0.34	0.13	0.45	0.81	0.45

Results in women are provided in Supplemental Table 1. T and SHBG concentrations are given in nanomoles per liter, and the estrogens are given in picomoles per liter. Sex hormones ranged across quartiles 1–4 for TT are <14.5, 14.5/19.5, 19.5/24.8, and >24.8; for free T are <0.23, 0.23/0.29, 0.29/0.35, and >0.35; for SHBG are <39.7, 39.7/53.3, 53.3/72.6, and >72.6; for E1 are <144.04, 144.04/184.25, 184.25/223.9, and >223.9; and for E2 are <73.97, 73.97/92.88, 92.88/115.9, and >115.9. To convert the values of TT from nanomoles per liter to nanograms per deciliter, multiply by 28.82. To convert the values for free T to picograms per milliliter, divide by 3.467. To convert E2 and E1 from picomoles per liter to picograms per milliliter, divide the E2 value by 3.671 and E1 by 3.699. Clinical PAD was defined as a composite outcome including an ABI below 0.90, IC, or lower extremity revascularization. Number of men exhibiting clinical PAD at baseline was 89 (6.3%). Ref., Reference.

[a] Results are obtained from multiple linear (ABI) and logistic regression (PAD) with statistical control for the confounding influences of age, waist circumference, smoking status, total and high-density lipoprotein cholesterol, diabetes, hypertension, and cardiovascular disease. Estimates are interpreted as differences or relative odds expressed vis-a-vis the reference sex hormone quartile.

[b] $P < 0.05$.

at 18.5 nmol/liter (532 ng/dl), and 3.3% at 35.3 nmol/liter (1016 ng/dl), respectively. E2 and E1 were measured simultaneously by LC-MS/MS with functional sensitivity of 2.5 pg/ml with use of volume dilution for samples with very low levels (24). For E2, the interassay CV were 9.4% at 29.6 pg/ml, 8.9% at 57.6 pg/ml, 7.6% at 109.9 pg/ml, and 7.6% at 329.3 pg/ml. The interassay CV for E1 were 13.5% at 29.8 pg/ml, 13.8% at 58.8 pg/ml, 12.4% for 114.2 pg/ml, and 11.4% for 341.0 pg/ml. SHBG concentrations were measured using an immunofluorometric assay (DELFIA-Wallac, Inc., Turku, Finland). The interassay CV were 8.3, 7.9, and 10.9% in the low, medium, and high pools (26, 27). Free T was calculated by using a modified law of mass action equation (28).

Peripheral arterial disease

Ankle-brachial index

Ankle-brachial systolic blood pressure measurements were obtained in the seventh and eighth examination cycles of the FHS by trained technicians according to a standard protocol, as previously described (29). Systolic blood pressure was measured twice in both arms and both ankles using an 8-MHz Doppler pen probe and an ultrasonic Doppler flow detector (Parks Medical Electronics, Inc., Aloha, OR). A third measurement was taken when the initial and second blood pressure measurements differed by more than 10 mm Hg at any site. Measurements were obtained from the dorsalis pedis artery only if the posterior tibial pulse could not be located by palpation or with Doppler probe. ABI was calculated as the ratio of the mean systolic blood pressure in the ankle divided by the systolic blood pressure in the arm with the greater mean. The lower of the two ABI ratios was used for analysis (29). A clinically meaningful change in ABI was defined as a decline of at least 0.15 between baseline and follow-up (30, 31) because this level of ABI decline is associated with a significant increase in risk for CVD (31).

Composite clinical PAD

Intermittent claudication (IC) was assessed using a standardized physician-administered questionnaire that inquired about the presence of exertional calf discomfort related to walking uphill or walking rapidly and was relieved with rest. Two physicians independently interviewed all participants with responses indicative of IC. An endpoint panel, comprised of three senior investigators, examined all medical evidence and made the final diagnosis of the presence of IC. Participants were also queried about revascularization procedures including lower extremity bypass surgery and percutaneous transluminal angioplasty. The endpoint panel used hospital records to validate all cardiovascular procedures, and the date and type of procedure was recorded. PAD was defined as a composite outcome including an ABI below 0.90, IC, or lower extremity revascularization.

Covariates

Sociodemographic and behavioral characteristics as well as medical history and medication use were assessed by standardized personal interviews. Current smokers were defined as those who reported having smoked at least one cigarette per day regularly during the year preceding the exam. Waist circumference, height, and weight were measured with the subject standing and body mass index calculated (kilograms per square meter). The examining physician measured resting blood pressure twice, and the average of the two readings was used to determine the presence of hypertension. Hypertension was defined as systolic blood pressure of at least 140 mm Hg or diastolic blood pressure of at least 90 mm Hg or self-reported use of antihypertensive medication. Diabetes was defined as fasting glucose of at least 126 mg/dl or self-reported use of insulin or oral hypoglycemic medications. Fasting plasma total cholesterol and high-density lipoprotein cholesterol concentrations were measured using standard enzymatic methods, as previously described (32). Prevalent

TABLE 2. Continued

E1		E2		TT/E1		TT/E2	
ABI, β (95% CI)	PAD, OR (95% CI)	ABI, β (95% CI)	PAD, OR (95% CI)	ABI, β (95% CI)	PAD, OR (95% CI)	ABI, β (95% CI)	PAD, OR (95% CI)
Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
−0.01 (−0.03, 0.01)	1.57 (0.80, 3.07)	−0.01 (−0.03, 0.01)	0.79 (0.41, 1.51)	0.02 (0.001, 0.04)[b]	0.64 (0.37, 1.13)	0.01 (−0.003, 0.03)	0.81 (0.47, 1.41)
−0.002 (−0.02, 0.02)	1.34 (0.67, 2.67)	−0.004 (−0.02, 0.01)	1.01 (0.55, 1.85)	0.03 (0.01, 0.05)[b]	0.62 (0.35, 1.11)	0.02 (0.004, 0.04)[b]	0.50 (0.26, 0.95)[b]
−0.03 (−0.05, −0.01)[b]	1.64 (0.85, 3.17)	−0.008 (−0.03, 0.01)	0.85 (0.45, 1.61)	0.03 (0.02, 0.05)[b]	0.34 (0.17, 0.69)[b]	0.03 (0.01, 0.05)[b]	0.52 (0.27, 1.01)
0.002	0.47	0.60	0.84	0.001	0.02	0.01	0.09
Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
−0.007 (−0.02, 0.01)	1.52 (0.71, 3.28)	−0.01 (−0.03, 0.01)	0.84 (0.40, 1.75)	0.01 (−0.01, 0.03)	0.71 (0.38, 1.32)	0.01 (−0.01, 0.03)	1.12 (0.60, 2.09)
0.004 (−0.01, 0.02)	1.32 (0.61, 2.88)	−0.002 (−0.02, 0.02)	1.05 (0.53, 2.08)	0.02 (0.003, 0.04)[b]	0.89 (0.46, 1.75)	0.02 (0.001, 0.04)[b]	0.71 (0.34, 1.48)
−0.02 (−0.04, −0.002)[b]	1.30 (0.62, 2.76)	−0.006 (−0.02, 0.01)	0.85 (0.42, 1.72)	0.02 (0.01, 0.04)[b]	0.47 (0.21, 1.05)	0.02 (0.004, 0.04)[b]	0.79 (0.36, 1.71)
0.05	0.76	0.55	0.89	0.05	0.27	0.10	0.60

CVD was assessed according to previously reported standardized protocols [including coronary heart disease (recognized or unrecognized myocardial infarction, angina pectoris, coronary insufficiency, or coronary heart disease death), cerebrovascular disease (stroke or transient ischemic attack), or congestive heart failure] and confirmed with the aid of medical histories, physical examinations at the study clinic, and hospitalization records and, finally, validated by the endpoint committee (22).

Statistical analyses

Sex hormone concentrations were categorized into quartiles. Given the known gender dimorphism in sex hormone effects, models were sex specific. In women, models were additionally stratified by menopausal status and postmenopausal use of hormone therapy. Independent analyses were performed for each sex hormone. Exploratory models included graphical and tabular displays to determine patterns of association and magnitude of change in ABI between examinations. Multivariable linear and logistic regression models adjusted for age, waist circumference, smoking status, total and high-density lipoprotein cholesterol, diabetes, hypertension, and prevalent cardiovascular disease were used, with estimates expressed as linear regression coefficient (β) or odds ratio (OR) and their corresponding 95% confidence interval (CI). Longitudinal associations of baseline sex hormone concentrations with meaningful change in ABI (decline of $\geq$0.15 between baseline and follow-up) were analyzed only among individuals without baseline PAD using multivariable logistic regression models. A separate sensitivity analysis assessed the mediating effects of statin use. Two-sided P values <0.05 were considered statistically significant. All statistical analyses were performed using SAS statistical software (SAS Institute Inc., Cary, NC).

Results

Table 1 presents the baseline characteristics of the study population according to sex, menopausal status, and postmenopausal use of hormone therapy. Postmenopausal women, mean age 64.6 yr, constituted the majority of our female study sample (86.5%). Clinical PAD was prevalent among 6.3% (n = 89) of men, 1.4% (n = 3) of premenopausal women, 3.7% (n = 19) of postmenopausal women using hormone therapy, and 6.7% (n = 59) of postmenopausal women not using hormone therapy.

Cross-sectional multivariable linear regression models revealed that men with lower free T and higher E1 concentrations had a significantly lower ABI (for free T, lowest *vs.* higher quartiles, $\beta = -0.02$, with 95% CI = -0.04 to -0.01; and for E1, highest *vs.* lower quartiles, $\beta = -0.03$, with 95% CI = -0.05 to -0.01, respectively). A higher TT/E1 and TT/E2 ratio was associated with higher ABI after adjustment for age and other confounders (Table 2). Furthermore, lower TT and SHBG concentrations in men were associated with prevalent PAD in age-adjusted (OR = 2.24, with 95% CI = 1.17–4.32; and OR = 2.06, with 95% CI = 1.07–3.96, lowest *vs.* highest quartile, respectively, although there was no significant trend noted across TT, $P = 0.08$, and SHBG, $P = 0.17$, quartiles) but not in multivariable logistic regression models (Table 2). Similarly, a higher TT/E1 ratio showed a protective effect on prevalent PAD in age-adjusted (OR = 0.34; 95% CI = 0.17–0.69) but not in multivariable-adjusted logistic regression models (Table 2). In women, we found no cross-sectional associations of any sex hormone with ABI (Supplemental Table 1, published on The Endocrine Society's Journals Online web site at http://jcem.endojournals.org), and the number of prevalent PAD cases (overall n = 81) was too low to yield meaningful estimates when stratifying by menopause and postmenopausal use of hormone therapy (data not shown).

Over the 6.7-yr (median) follow-up period, the distribution of measured ABI values in men showed no statistically significant differences (Student's t test, $P = 0.78$), although a substantial number of men (n = 69) experienced a meaningful change in ABI (decline of $\geq$0.15) between baseline and follow-up (Fig. 2 and Supplemental Fig. 1). Multivariable logistic regression models revealed an association of baseline SHBG concentrations with meaningful change in ABI among men (n = 1076 men free of baseline PAD and data at exam 8). OR for men in SHBG quartiles 1, 2, and 3 compared with quartile 4 were 2.56 (95% CI = 1.01–6.45), 2.28 (95% CI = 0.98–5.32), and 2.93 (95% CI = 1.31–6.52), respectively (P for trend = 0.07). No such association was observed for TT, free T, E1, E2, or TT/E ratios (Table 3). The number of incident PAD cases was too low (women, n = 24; men, n = 21) to perform incidence analyses (Supplemental Fig. 1). Also the number of women with meaningful change in ABI was too low (n = 6 premenopausal women; n = 24 postmenopausal women on HT; n = 43 postmenopausal women not on HT) to perform any regression modeling. Additional adjustment for statin use showed no impact on the overall estimates (data not shown).

Discussion

Principal findings

The present study is the first investigation of cross-sectional and longitudinal associations between sex hormone concentrations with ABI, PAD, and ABI change, analyzing data from a large community-based sample of men and women. In men, cross-sectional analyses revealed a positive association of free T and a negative association of E1 concentrations with ABI, respectively. Longitudinal analyses also showed sex-specific associations of SHBG concentrations with ABI change in men.

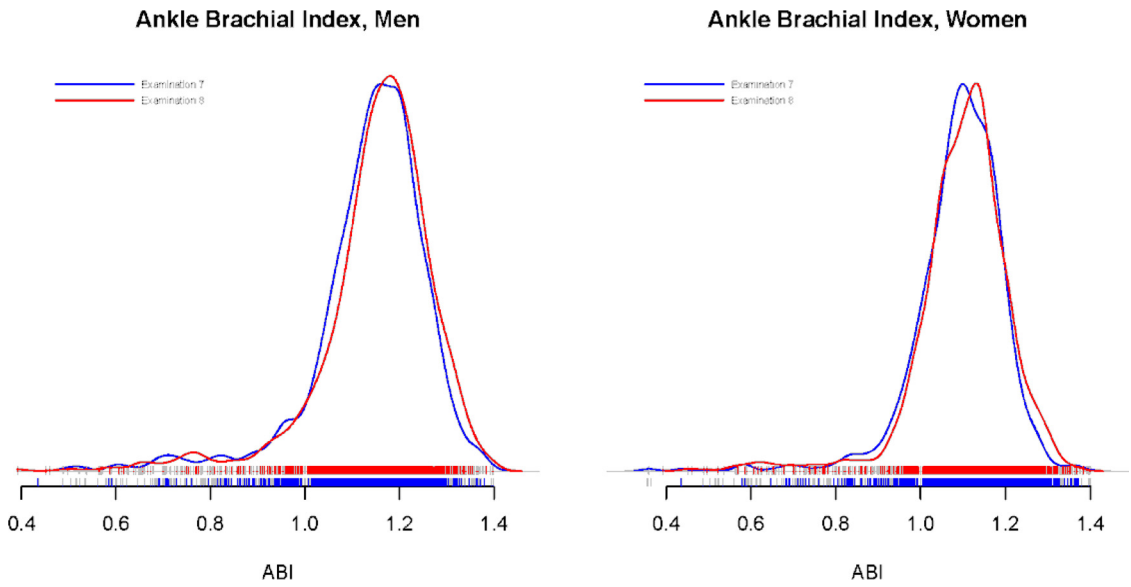


FIG. 2. Kernel density estimates for ABI, by sex and examination. *Hash marks at bottom* depict actual ABI values.

Findings in men

Previous studies examining the association of sex hormones with PAD among men are very sparse. One previous cross-sectional study among 3014 elderly men (mean age 74.5 yr) found that low free T and TT concentrations were associated with ABI and prevalent PAD (defined as an ABI <0.90) (21). But in contrast to our results, this previous study showed a significant association between TT and prevalent PAD after multivariable adjustment, suggesting that differences in the adjustment set could have mitigated or abolished significant effects of TT on PAD in our analyses. The previous study also differed from our study in that no evidence of an association between SHBG and PAD or ABI was observed (21).

Although E1 is at least as abundant in circulation as E2, its biological role remains poorly understood. We found a negative association between E1 concentrations and ABI in cross-sectional models. However, in line with previous findings, E2 was not associated with ABI or prevalent PAD in our study (21). The potential role of sex hormones in vascular disease has been reinforced by a recent cross-sectional study among 3620 community-dwelling men aged 70–88 yr, showing that low free T and high SHBG concentrations are independently associated with abdominal aortic aneurysm, itself a powerful predictor of cardiovascular events and mortality (33). Additionally, previous studies reported associations between low free T and TT concentrations with carotid plaques and carotid in-

TABLE 3. Longitudinal associations of sex hormone concentrations in men with meaningful change (decline of at least 0.15) in ABI

	OR (95% CI)						
	TT	Free T	SHBG	E1	E2	TT/E1	TT/E2
Age-adjusted models							
<25th	1.97 (0.98, 3.97)	1.87 (0.85, 4.14)	2.54 (1.09, 5.94)	Ref.	Ref.	Ref.	Ref.
25–50th	1.4 (0.67, 2.94)	1.18 (0.51, 2.73)	2.35 (1.05, 5.25)	1.60 (0.73, 3.48)	1.16 (0.58, 2.34)	0.98 (0.53, 1.81)	1.02 (0.54, 1.92)
50–75th	0.97 (0.44, 2.14)	1.85 (0.85, 4.03)	2.66 (1.23, 5.75)	1.89 (0.88, 4.06)	0.99 (0.49, 2.02)	0.54 (0.26, 1.10)	0.82 (0.41, 1.65)
>75th	Ref.	Ref.	Ref.	1.89 (0.88, 4.06)	1.04 (0.51, 2.13)	0.41 (0.19, 0.90)	0.50 (0.23, 1.09)
P for Trend	0.14	0.25	0.08	0.35	0.97	0.05	0.28
Multivariable-adjusted models[a]							
<25th	1.89 (0.85, 4.20)	1.56 (0.68, 3.59)	2.56 (1.01, 6.45)[b]	Ref.	Ref.	Ref.	Ref.
25–50th	1.37 (0.63, 3.01)	1.02 (0.43, 2.42)	2.28 (0.98, 5.32)	1.51 (0.67, 3.36)	1.32 (0.64, 2.73)	0.99 (0.51, 1.91)	1.10 (0.56, 2.16)
50–75th	0.99 (0.44, 2.27)	1.63 (0.73, 3.62)	2.93 (1.31, 6.52)[b]	1.78 (0.81, 3.88)	1.01 (0.48, 2.14)	0.64 (0.30, 1.36)	0.87 (0.40, 1.87)
>75th	Ref.	Ref.	Ref.	1.49 (0.67, 3.29)	1.04 (0.49, 2.19)	0.45 (0.19, 1.05)	0.54 (0.22, 1.30)
P for trend	0.30	0.43	0.07	0.55	0.84	0.18	0.38

For men free of clinical PAD at examination 7 and ABI data at examination 8, n = 1076. For meaningful change in ABI (decline of at least 0.15), n = 69. Ref., Reference.

[a] Estimates are adjusted for age, waist circumference, smoking status, total and high-density lipoprotein cholesterol, diabetes, hypertension, and cardiovascular disease.

[b] $P < 0.05$.

tima media thickness, respectively (34–36). Taken together, present evidence about the potential role of sex hormones in vascular disease is largely based on cross-sectional studies, requiring future prospective studies to dissect the temporal relationship between sex hormones and the atherosclerotic process.

Thus, we conducted longitudinal analyses to investigate the effect of baseline T, estrogen, and SHBG concentrations on ABI change. We found that SHBG concentrations in men were associated with meaningful ABI decline ($\geq$0.15) between baseline and follow-up, but without any evidence for a linear trend across SHBG quartiles. Furthermore, the comparably large CI reflect a lack of precision for the revealed risk estimates due to the low number of men with meaningful change in ABI. Similarly, we were not able to investigate longitudinal associations between sex hormone concentrations and incident PAD because the number of incident cases was too low. However, preliminary results from a small cross-sectional case-control study among PAD patients showed no significant associations between sex hormone concentrations and clinical PAD in either sex (37). Furthermore, associations between sex hormone concentrations and incident clinical CVD are conflicting (38, 39). Therefore, the extent to which SHBG and other sex hormones could be a useful risk marker or exert independent effects on the atherosclerosis process needs to be determined.

Findings in women

In the present study, we did not observe any associations of sex hormone concentrations with ABI or PAD in women. It is also possible that sex hormones play a role in the atherosclerotic process earlier in life, independent of the substantial decrease in estrogen concentrations during the menopause. Therefore, we performed stratified analyses with regard to menopausal status and postmenopausal use of hormone therapy but without revealing any statistically significant differences between these groups. We also considered postmenopausal hormone therapy (*i.e.* exogenous estrogen or estrogen/progestin combinations) as a potential mediator of the investigated associations but were not able to detect different associations of sex hormones with ABI or PAD in postmenopausal women who used hormone therapy *vs.* those who did not. Similarly, previous studies of hormone therapy in postmenopausal women showed no effect of E2 on progression of carotid atherosclerosis (40, 41). However, the lack of associations observed in our study may reflect a lack of statistical power to detect differential effects of sex hormones in women within menopause and hormone therapy subgroups.

Previous studies investigating the relation between sex hormone concentrations and cardiovascular risk factors in women revealed associations of higher T (free T and TT) and lower SHBG concentrations with increased cardiovascular risk factor burden (12) including visceral fat accumulation (13), insulin resistance, adverse lipid profiles (14), diabetes (15), and incident CVD (19, 20). In particular, higher SHBG concentrations were associated with reduced subclinical atherosclerosis progression in healthy postmenopausal women (16–18). In contrast, two studies have shown that lower circulating TT may be associated with incident cardiovascular morbidity (42) and heart disease (43) but were limited by imprecise immunoassay-based TT measurements in the low concentration range of women and missing free or bioavailable T assessment. Given the conflicting reports in the literature, firm conclusions about the association between sex hormones and clinical CVD events in women cannot be drawn (12). However, because our power to detect an association was low, the present lack of statistically significant association of sex hormones with ABI and PAD in women needs to be interpreted with caution until further investigation in other large prospective epidemiological studies with long-term follow-up.

To determine whether exogenous sex hormones are an effective treatment for PAD patients, a review identified only three, small-scale trials in predominately male populations, involving the use of T (no estrogenic hormones) over relatively short time periods with variable methods of measuring PAD (44). Overall, T supplementation had no significant effect on PAD progression, including subjective improvement in symptoms and tests of walking distance (44). Additionally, a recent trial among elderly men reported that T supplementation increased the risk of adverse cardiovascular events (45). However, it is still possible that exogenous T could affect PAD progression, but at present, there are insufficient data to support this hypothesis.

Potential limitations of the present study include the small number of PAD events precluding incidence analyses and constraining analyses of ABI change as well as a study sample of adult white men and women limiting the generalizability of our findings to individuals of other ethnicities. Strengths of the present study include its large community-based sample; TT, E2, and E1 concentrations measured by LC-MS/MS from fasting morning samples; and a comprehensive directly measured covariate assessment.

In conclusion, our investigation of a middle-aged community-based sample suggests that lower free T and higher E1 concentrations may be associated with PAD and ABI change in men, but sex hormones were not associated with

PAD or ABI in women. Additional investigations are warranted to confirm these findings and to elucidate the biological basis for the sex-related differences in associations between sex hormones and PAD.

Acknowledgments

Address all correspondence and requests for reprints to: Mr. Robin Haring, Ph.D., University of Greifswald Institute of Clinical Chemistry and Laboratory Medicine, Ferdinand-Sauerbruch-Strasse, 17475 Greifswald, Germany. E-mail: robin.haring@uni-greifswald.de.

The Framingham Heart Study is funded by National Heart, Blood, and Lung Institute (contract N01-HC-25195). Additional support was provided by National Institutes of Health Grant 1RO1AG31206 (to S.B. and R.S.V.) and RO1HL094755-01A2 (to A.D.C., V.R., and S.B.). R.H. was supported by a scholarship from the Alfried Krupp von Bohlen und Halbach Foundation.

Disclosure Summary: R.H., T.G.T., R.S.V., M.N.D., A.D., and J.M.M. have nothing to disclose. S.B. has received research grants from Solvay Pharmaceuticals, Merck and Co., and Ligand Pharmaceuticals for research unrelated to the contents of this manuscript. S.B. serves as a member of the American Board of Internal Medicine, Endocrinology, and Metabolism and chaired The Endocrine Society's Expert Panel for the development of guidelines for Testosterone Therapy of Androgen Deficiency Syndromes in Men. H.W. has received research grants from Bayer, Novartis, Novo Nordisk, and Pfizer for research unrelated to the contents of this manuscript and serves as a member of the International and German KIMS board.

References

1. **Hirsch AT, Haskal ZJ, Hertzer NR, Bakal CW, Creager MA, Halperin JL, Hiratzka LF, Murphy WR, Olin JW, Puschett JB, Rosenfield KA, Sacks D, Stanley JC, Taylor Jr LM, White CJ, White J, White RA, Antman EM, Smith Jr SC, Adams CD, Anderson JL, Faxon DP, Fuster V, Gibbons RJ, Halperin JL, *et al.*** 2006 ACC/AHA 2005 guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric, and abdominal aortic): executive summary a collaborative report from the American Association for Vascular Surgery/Society for Vascular Surgery, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, Society of Interventional Radiology, and the ACC/AHA Task Force on Practice Guidelines (Writing Committee to Develop Guidelines for the Management of Patients With Peripheral Arterial Disease) endorsed by the American Association of Cardiovascular and Pulmonary Rehabilitation; National Heart, Lung, and Blood Institute; Society for Vascular Nursing; TransAtlantic Inter-Society Consensus; and Vascular Disease Foundation. J Am Coll Cardiol 47:1239–1312
2. **Fowkes FG, Murray GD, Butcher I, Heald CL, Lee RJ, Chambless LE, Folsom AR, Hirsch AT, Dramaix M, deBacker G, Wautrecht JC, Kornitzer M, Newman AB, Cushman M, Sutton-Tyrrell K, Fowkes FG, Lee AJ, Price JF, d'Agostino RB, Murabito JM, Norman PE, Jamrozik K, Curb JD, Masaki KH, Rodríguez BL, *et al.*** 2008 Ankle brachial index combined with Framingham Risk Score to predict cardiovascular events and mortality: a meta-analysis. JAMA 300:197–208
3. **Resnick HE, Lindsay RS, McDermott MM, Devereux RB, Jones KL, Fabsitz RR, Howard BV** 2004 Relationship of high and low ankle brachial index to all-cause and cardiovascular disease mortality: the Strong Heart Study. Circulation 109:733–739
4. **Criqui MH, Langer RD, Fronek A, Feigelson HS, Klauber MR, McCann TJ, Browner D** 1992 Mortality over a period of 10 years in patients with peripheral arterial disease. N Engl J Med 326:381–386
5. **Heald CL, Fowkes FG, Murray GD, Price JF** 2006 Risk of mortality and cardiovascular disease associated with the ankle-brachial index: systematic review. Atherosclerosis 189:61–69
6. **Khaw KT, Barrett-Connor E** 1992 Lower endogenous androgens predict central adiposity in men. Ann Epidemiol 2:675–682
7. **Haring R, Völzke H, Felix SB, Schipf S, Dörr M, Rosskopf D, Nauck M, Schöfl C, Wallaschofski H** 2009 Prediction of metabolic syndrome by low serum testosterone levels in men: results from the study of health in Pomerania. Diabetes 58:2027–2031
8. **Vikan T, Schirmer H, Njølstad I, Svartberg J** 2010 Low testosterone and sex hormone-binding globulin levels and high estradiol levels are independent predictors of type 2 diabetes in men. Eur J Endocrinol 162:747–754
9. **Haring R, Baumeister SE, Völzke H, Dörr M, Felix SB, Kroemer HK, Nauck M, Wallaschofski H** 2011 Prospective association of low total testosterone concentrations with an adverse lipid profile and increased incident dyslipidemia. Eur J Cardiovasc Prev Rehabil 18:86–96
10. **Liu PY, Death AK, Handelsman DJ** 2003 Androgens and cardiovascular disease. Endocr Rev 24:313–340
11. **Haring R, Völzke H, Steveling A, Krebs A, Felix SB, Schöfl C, Dörr M, Nauck M, Wallaschofski H** 2010 Low serum testosterone levels are associated with increased risk of mortality in a population-based cohort of men aged 20–79. Eur Heart J 31:1494–1501
12. **Brand JS, van der Schouw YT** 2010 Testosterone, SHBG and cardiovascular health in postmenopausal women. Int J Impot Res 22:91–104
13. **Janssen I, Powell LH, Kazlauskaite R, Dugan SA** 2010 Testosterone and visceral fat in midlife women: the Study of Women's Health Across the Nation (SWAN) fat patterning study. Obesity (Silver Spring) 18:604–610
14. **Sutton-Tyrrell K, Wildman RP, Matthews KA, Chae C, Lasley BL, Brockwell S, Pasternak RC, Lloyd-Jones D, Sowers MF, Torréns JI** 2005 Sex-hormone-binding globulin and the free androgen index are related to cardiovascular risk factors in multiethnic premenopausal and perimenopausal women enrolled in the Study of Women Across the Nation (SWAN). Circulation 111:1242–1249
15. **Golden SH, Dobs AS, Vaidya D, Szklo M, Gapstur S, Kopp P, Liu K, Ouyang P** 2007 Endogenous sex hormones and glucose tolerance status in postmenopausal women. J Clin Endocrinol Metab 92:1289–1295
16. **Ouyang P, Vaidya D, Dobs A, Golden SH, Szklo M, Heckbert SR, Kopp P, Gapstur SM** 2009 Sex hormone levels and subclinical atherosclerosis in postmenopausal women: the Multi-Ethnic Study of Atherosclerosis. Atherosclerosis 204:255–261
17. **Karim R, Hodis HN, Stanczyk FZ, Lobo RA, Mack WJ** 2008 Relationship between serum levels of sex hormones and progression of subclinical atherosclerosis in postmenopausal women. J Clin Endocrinol Metab 93:131–138
18. **Calderon-Margalit R, Schwartz SM, Wellons MF, Lewis CE, Daviglus ML, Schreiner PJ, Williams OD, Sternfeld B, Carr JJ, O'Leary DH, Sidney S, Friedlander Y, Siscovick DS** 2010 Prospective association of serum androgens and sex hormone-binding globulin with subclinical cardiovascular disease in young adult women: the "Coronary Artery Risk Development in Young Adults" women's study. J Clin Endocrinol Metab 95:4424–4431
19. **Guthrie JR, Dennerstein L, Taffe JR, Lehert P, Burger HG** 2004 The menopausal transition: a 9-year prospective population-based

study. The Melbourne Women's Midlife Health Project. Climacteric 7:375–389
20. **Patel SM, Ratcliffe SJ, Reilly MP, Weinstein R, Bhasin S, Blackman MR, Cauley JA, Sutton-Tyrrell K, Robbins J, Fried LP, Cappola AR** 2009 Higher serum testosterone concentration in older women is associated with insulin resistance, metabolic syndrome, and cardiovascular disease. J Clin Endocrinol Metab 94:4776–4784
21. **Tivesten A, Mellström D, Jutberger H, Fagerberg B, Lernfelt B, Orwoll E, Karlsson MK, Ljunggren O, Ohlsson C** 2007 Low serum testosterone and high serum estradiol associate with lower extremity peripheral arterial disease in elderly men. The MrOS Study in Sweden. J Am Coll Cardiol 50:1070–1076
22. **Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP** 1979 An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol 110:281–290
23. **Bhasin S, Pencina M, Jasuja GK, Travison TG, Coviello A, Orwoll E, Wang PY, Nielson C, Wu F, Tajar A, Labrie F, Vesper H, Zhang A, Ulloor J, Singh R, D'Agostino R, Vasan RS** 2011 Reference ranges for testosterone in men generated using liquid chromatography tandem mass spectrometry in a community-based sample of healthy nonobese young men in the Framingham Heart Study and applied to three geographically distinct cohorts. J Clin Endocrinol Metab 96: 2430–2439
24. **Nelson RE, Grebe SK, OKane DJ, Singh RJ** 2004 Liquid chromatography-tandem mass spectrometry assay for simultaneous measurement of estradiol and estrone in human plasma. Clin Chem 50:373–384
25. **Vesper HW, Bhasin S, Wang C, Tai SS, Dodge LA, Singh RJ, Nelson J, Ohorodnik S, Clarke NJ, Salameh WA, Parker Jr CR, Razdan R, Monsell EA, Myers GL** 2009 Interlaboratory comparison study of serum total testosterone [corrected] measurements performed by mass spectrometry methods. Steroids [Erratum (2009) 74:791] 74: 498–503
26. **Bhasin S, Storer TW, Javanbakht M, Berman N, Yarasheski KE, Phillips J, Dike M, Sinha-Hikim I, Shen R, Hays RD, Beall G** 2000 Testosterone replacement and resistance exercise in HIV-infected men with weight loss and low testosterone levels. JAMA 283:763–770
27. **Bhasin S, Woodhouse L, Casaburi R, Singh AB, Mac RP, Lee M, Yarasheski KE, Sinha-Hikim I, Dzekov C, Dzekov J, Magliano L, Storer TW** 2005 Older men are as responsive as young men to the anabolic effects of graded doses of testosterone on the skeletal muscle. J Clin Endocrinol Metab 90:678–688
28. **Mazer NA** 2009 A novel spreadsheet method for calculating the free serum concentrations of testosterone, dihydrotestosterone, estradiol, estrone and cortisol: with illustrative examples from male and female populations. Steroids 74:512–519
29. **Murabito JM, Evans JC, Nieto K, Larson MG, Levy D, Wilson PW** 2002 Prevalence and clinical correlates of peripheral arterial disease in the Framingham Offspring Study. Am Heart J 143:961–965
30. **Kennedy M, Solomon C, Manolio TA, Criqui MH, Newman AB, Polak JF, Burke GL, Enright P, Cushman M** 2005 Risk factors for declining ankle-brachial index in men and women 65 years or older: the Cardiovascular Health Study. Arch Intern Med 165:1896–1902
31. **Criqui MH, Ninomiya JK, Wingard DL, Ji M, Fronek A** 2008 Progression of peripheral arterial disease predicts cardiovascular disease morbidity and mortality. J Am Coll Cardiol 52:1736–1742
32. **McNamara JR, Schaefer EJ** 1987 Automated enzymatic standardized lipid analyses for plasma and lipoprotein fractions. Clin Chim Acta 166:1–8
33. **Yeap BB, Hyde Z, Norman PE, Chubb SA, Golledge J** 2010 Associations of total testosterone, sex hormone-binding globulin, calculated free testosterone, and luteinizing hormone with prevalence of abdominal aortic aneurysm in older men. J Clin Endocrinol Metab 95:1123–1130
34. **Dörr M, Wallaschofski H, Friedrich N** 2009 Association of low total testosterone levels and prevalent carotid plaques: result of the study of health in Pomerania. Eur J Epidemiol 24:389–391
35. **Vikan T, Johnsen SH, Schirmer H, Njølstad I, Svartberg J** 2009 Endogenous testosterone and the prospective association with carotid atherosclerosis in men: the Tromso study. Eur J Epidemiol 24:289–295
36. **De Pergola G, Pannacciulli N, Ciccone M, Tartagni M, Rizzon P, Giorgino R** 2003 Free testosterone plasma levels are negatively associated with the intima-media thickness of the common carotid artery in overweight and obese glucose-tolerant young adult men. Int J Obes Relat Metab Disord 27:803–807
37. **Price JF, Lee AJ, Fowkes FG** 1997 Steroid sex hormones and peripheral arterial disease in the Edinburgh Artery Study. Steroids 62:789–794
38. **Arnlöv J, Pencina MJ, Amin S, Nam BH, Benjamin EJ, Murabito JM, Wang TJ, Knapp PE, D'Agostino RB Sr, Bhasin S, Vasan RS** 2006 Endogenous sex hormones and cardiovascular disease incidence in men. Ann Intern Med 145:176–184
39. **Yeap BB, Hyde Z, Almeida OP, Norman PE, Chubb SA, Jamrozik K, Flicker L, Hankey GJ** 2009 Lower testosterone levels predict incident stroke and transient ischemic attack in older men. J Clin Endocrinol Metab 94:2353–2359
40. **Karim R, Mack WJ, Lobo RA, Hwang J, Liu CR, Liu CH, Sevanian A, Hodis HN** 2005 Determinants of the effect of estrogen on the progression of subclinical atherosclerosis: Estrogen in the Prevention of Atherosclerosis Trial. Menopause 12:366–373
41. **Byington RP, Furberg CD, Herrington DM, Herd JA, Hunninghake D, Lowery M, Riley W, Craven T, Chaput L, Ireland CC, Applegate WB** 2002 Effect of estrogen plus progestin on progression of carotid atherosclerosis in postmenopausal women with heart disease: HERS B-mode substudy. Arteriosclerosis, thrombosis, and vascular biology 22:1692–1697
42. **Sievers C, Klotsche J, Pieper L, Schneider HJ, März W, Wittchen HU, Stalla GK, Mantzoros C** 2010 Low testosterone levels predict all-cause mortality and cardiovascular events in women: a prospective cohort study in German primary care patients. Eur J Endocrinol 163:699–708
43. **Laughlin GA, Goodell V, Barrett-Connor E** 2010 Extremes of endogenous testosterone are associated with increased risk of incident coronary events in older women. J Clin Endocrinol Metab 95:740–747
44. **Price JF, Leng GC** 2000 Steroid sex hormones for lower limb atherosclerosis. Cochrane Database Syst Rev 2:CD000188
45. **Basaria S, Coviello AD, Travison TG, Storer TW, Farwell WR, Jette AM, Eder R, Tennstedt S, Ulloor J, Zhang A, Choong K, Lakshman KM, Mazer NA, Miciek R, Krasnoff J, Elmi A, Knapp PE, Brooks B, Appleman E, Aggarwal S, Bhasin G, Hede-Brierley L, Bhatia A, Collins L, LeBrasseur N, *et al.*** 2010 Adverse events associated with testosterone administration. N Engl J Med 363:109–122

Authors:

Bethlyn Vergo Houlihan, MSW, MPH
Alan Jette, PhD, PT
Michael Paasche-Orlow, MD, MA, MPH
Jane Wierbicky, RN
Stan Ducharme, PhD
Judi Zazula, MS, OTR/L
Penelope Cuevas, MD
Robert H. Friedman, MD
Steve Williams, MD

Affiliations:

From the Department of Health Policy & Management (BVH) and the Health & Disability Research Institute (AJ), Boston University School of Public Health; Section of General Internal Medicine, Department of Medicine (MP-O, PC, RHF), and the Department of Physical Medicine and Rehabilitation (SD, SW), Boston University School of Medicine; and the New England Regional Spinal Cord Injury Center (BVH, JW, JZ, SW) and Rehabilitation Medicine (SD), Boston Medical Center, Massachusetts.

Correspondence:

All correspondence and requests for reprints should be addressed to: Bethlyn Vergo Houlihan, MSW, MPH, New England Regional SCI Center, Boston Medical Center, 732 Harrison Ave, F511, Boston, MA 02118.

Disclosures:

Financial disclosure statements have been obtained, and no conflicts of interest have been reported by the authors or by any individuals in control of the content of this article. Supported by Centers for Disease Control and Prevention grant 5R01DD000155, Department of Health and Human Services. Presented in poster format (portions of content) at the Spinal Cord Injury Contemporary Forums Conference, Orlando, FL, 2009; the American Congress of Rehabilitation Medicine Annual Conference, Denver, CO, 2009; and the International Spinal Cord Society Annual Meeting, Florence, Italy, 2009.

0894-9115/11/9009-0756/0
American Journal of Physical Medicine & Rehabilitation

DOI: 10.1097/PHM.0b013e31820b140f

BRIEF REPORT

A Telerehabilitation Intervention for Persons with Spinal Cord Dysfunction

ABSTRACT

Houlihan BV, Jette A, Paasche-Orlow M, Wierbicky J, Ducharme S, Zazula J, Cuevas P, Friedman RH, Williams S: A telerehabilitation intervention for persons with spinal cord dysfunction. Am J Phys Med Rehabil 2011;90:756–764.

Pressure ulcers and depression are common preventable conditions secondary to a spinal cord dysfunction. However, few successful, low-cost preventive approaches have been identified. We have developed a dynamic automated telephone calling system, termed *Care Call*, to empower and motivate people with spinal cord dysfunction to improve their skin care, seek treatment for depression, and appropriately use the healthcare system. Herein, we describe the design and development of Care Call, its novel features, and promising preliminary results of our pilot testing. Voice quality testing showed that Care Call was able to understand all voice characteristics except very soft-spoken speech. Importantly, pilot study subjects felt Care Call could be particularly useful for people who are depressed, those with acute injury, and those without access to quality care. The results of a randomized controlled trial currently underway to evaluate Care Call will be available in 2011.

Key Words: Spinal Cord Injuries, Telemedicine, Pressure Ulcer, Depression

Pressure ulcers and depression are common preventable secondary conditions for people with spinal cord dysfunction (SCD). Unfortunately, diagnosis and/or treatment of secondary conditions is often delayed.[1–4] This can undermine rehabilitation and have a significant impact on a person's quality-of-life and his/her healthcare costs.[5–8]

A patient's self-care behavior can impact the onset and severity of secondary conditions after an SCD. Pressure ulcers, for example, can be prevented through the use of patient education.[9] Depression can be improved and successfully managed if a patient receives treatment.[10] Nonetheless, persons with SCD do not generally receive the necessary follow-up care[11] that could promote self-management behaviors. Patients report various obstacles to complying with self-care management and annual follow-up evaluations, including cost, transportation, time, and reluctance to seek treatment,[12] such that they are not being successfully encouraged to prevent or seek treatment for pressure ulcers and depression.

Interventions to promote healthy behaviors, which can in turn prevent secondary conditions, are typically resource intensive.[13,14] Telerehabilitation and

related technologies are a promising strategy that, if successful, could improve the quality and reduce the cost of secondary prevention.[15] For persons with spinal cord injury (SCI), particular attention in telerehabilitation interventions has been paid to using telephone contact and video monitoring for the prevention and management of pressure ulcers.[16,17] However, because telerehabilitation interventions have not been evaluated, few successful, low-cost approaches that prevent secondary conditions have been identified[18,19]; such a system could bring not only substantial long-term cost savings but also enhanced quality-of-life of people with SCD.

With this goal in mind, we have developed *Care Call*—an innovative telerehabilitation intervention system designed to empower and motivate people with SCD to improve their skin care and mental health. The system does not replace face-to-face health care, rather, it supplements a clinician's role in long-term management after SCD. Care Call has the potential to help significant numbers of people after an SCD using a low-risk, low-cost approach that could be used for long-term patient monitoring and service provision. If efficacious, this intervention could be offered by clinicians to patients across multiple settings. Herein, we describe the design, development, and initial pilot testing of the Care Call intervention, with the ultimate goal of informing the final intervention protocol for an initial randomized controlled trial (RCT) now underway.

METHODS

Description of the Care Call Technology

The Care Call intervention is delivered from the Telephone-Linked Computer System (TLC),[20] an automated, interactive conversation system that speaks with a digitized human voice.[21] TLC, which functions as an at-home monitor, educator, and counselor for reinforcing or changing health-related behaviors, has been used to screen and monitor numerous diseases[22–27] and has been applied to important health-related behaviors.[28] Clinical trials show TLC to improve medication adherence,[29] increase exercise among the general[13] and elderly populations,[30] decrease the degree of dyspnea for people with chronic obstructive pulmonary disease,[31] and improve eating habits[32] and lower serum cholesterol levels through dietary changes.[31]

TLC uses an interactive voice response system to generate digitized speech over the telephone, a speech recognition software, a conversation control system that directs the content and flow of individual TLC conversations with users, and a database management system for storing user information and call logs. It is a call-in and call-out system; that is, users can call (from any telephone) at any time and the system will also call the user according to an adaptive scheduling protocol that reflects how well a person is doing with his/her self-care. If the system initiates a call and no contact is made, it will leave a message for the person to call TLC and will call again according to a call schedule protocol. TLC automatically produces reports on utilization statistics to assist in system operation, intervention evaluation, and patient monitoring.

Design of the Care Call Intervention

The target population for the Care Call intervention consists of persons with SCD who use a wheelchair at least 6 hrs a day. For the Care Call clinical trial, further exclusion criteria were developed, including having nontraumatic SCI diagnoses with fast progression (amyotrophic lateral sclerosis, postpolio, and metastatic disease of the spine), having severe major depression, and having a stage III or greater pressure ulcer. In practice, the appropriateness of Care Call for subgroups meeting these exclusion criteria would need to be evaluated on an individual basis.

We designed the Care Call intervention to (1) screen for pressure ulcers and depressive symptoms, (2) educate about the prevention of depression and pressure ulcers and the appropriate use of heath care services, and (3) alert a nurse telerehabilitation coordinator (NTC), when appropriate, for direct medical or mental health attention. Furthermore, we hypothesized a secondary goal, that Care Call would improve community integration and quality-of-life. Each of these goals is being evaluated in the RCT. An interdisciplinary team of rehabilitation professionals developed the content, design, and overall functionality of the Care Call system.

The content of the Care Call intervention is based on the Transtheoretical Model[33] and Social Cognitive Theory,[34] as well as on the heuristics of experienced counselors. The Transtheoretical Model posits that people at different stages of readiness to make a desired behavioral change will respond to different counseling messages. A fundamental precept in the Social Cognitive Theory is that individuals use self-referent thought to mediate between knowledge and behavior, which allows them to evaluate their own experiences and thought processes.[35] Through the process of self-evaluation,

individuals may alter their own thinking about their abilities and the outcomes of their actions and subsequently alter their behavior.

Care Call targets pressure ulcers and depression, in particular, because of their prominence for people with SCD and their preventability through self-management. The prevalence of pressure ulcers in a community-based sample of individuals with SCD is estimated at 33%.[36] Patient education is generally considered paramount for proper prevention of pressure ulcers.[36] The percentage of adults with disabilities who reported that feelings of depression kept them from being active was three times that of the general population, 28% compared with 7%, respectively.[15] Research has shown that TLC technology has been able to successfully change health behavior and disease outcomes.[13,20,23,32,37] Although there is little research on using TLC to prevent or treat depression, screening for another sensitive mental health issue, substance use, has proven effective using TLC.[25] In addition, several randomized trials over the last decade have demonstrated that telehealth interventions can decrease depression severity.[38,39]

General Description of Care Call

The Care Call scripts are organized into modules and integrate information relative to three targeted areas: skin care, depression and wellness, and healthcare utilization. Although most content is delivered by computer voice, the system also has recorded vignettes from people with SCD and recorded comments from healthcare professionals. Throughout each module, users are referred to local community and informational resources via the Care Call Resource Book (see below).

The Skin Care Module assesses and monitors old and new skin problems, trains users in skin care, and assesses and monitors risk factors such as incontinence and equipment needs. Care Call reviews barriers to skin care adherence to offer specific encouragement and advice.[34] To provide individualized messages, the system is adapted based on data collected at baseline (e.g., history of pressure ulcers, consistency of sensation, and level of paralysis) and on responses in previous calls.

The Depression and Wellness Module screens, monitors, and educates users on depression and adjustment. For those with existing untreated depression, Care Call provides education and encouragement to improve their understanding and management of depression. The system assesses satisfaction with treatment and promotes adherence. Care Call also has a wellness track for decreasing depression in which participants are assessed; educated about exercise, sleep habits, and alcohol use; and offered a brief relaxation exercise that can be done at any time.

As part of the depression track, a "Self-harm Protocol" was developed for use when a subject endorsed thoughts of hurting himself/herself within the past 2 wks. When used, a follow-up script would be immediately implemented (by Care Call, the NTC, or field staff) to assess a subject's risk of self-harm. If the subject was found to be at immediate risk of self-harm, the system or Care Call staff would page an on-call clinical psychologist with the subject's status. The clinical psychologist would then reply within 1 hr to assess the situation and contract for safety, when appropriate. Otherwise, subjects would be counseled that if they ever began to feel like they could hurt themselves, they should go to the emergency department, and they would be provided with a national hotline number. In all cases, the NTC follows up within 48 hrs to see how a subject is doing.

The Health Care Utilization Module tracks each person's medical and mental health appointments. Because seeking treatment to prevent these problems is paramount, Care Call both reviews with users logistical factors before a healthcare visit and coaches them in communicating with a provider during a visit. If users miss an appointment, Care Call assesses barriers and offers recommendations to overcome them.

Care Call integrates audiotaped vignettes from actual patients with SCD. We recorded interviews with nine key informants with SCD to illustrate tips and personal experiences on all three modular topics. The key informants represented a range of SCDs (including multiple sclerosis and SCI) and ethnic/racial backgrounds (seven white non-Hispanic, two black; five men, four women). Users hear at least one audio clip per call, with relevant audio clips embedded throughout. Additional vignettes provide advice from SCD clinicians.

Lastly, Care Call offers users the opportunity throughout to speak with the NTC and alerts the NTC to contact the user for follow-up on important issues that are identified. The main role of the NTC is to respond to Care Call alerts in a timely manner, providing appropriate referral, resources, and/or action steps for users. Care Call does not act as an emergency responder to medical situations or dispense medical advice; the NTC provides an important triage role to further ascertain the needs of each Care Call user beyond the limitations of automated technology. The NTC uses a Web-based

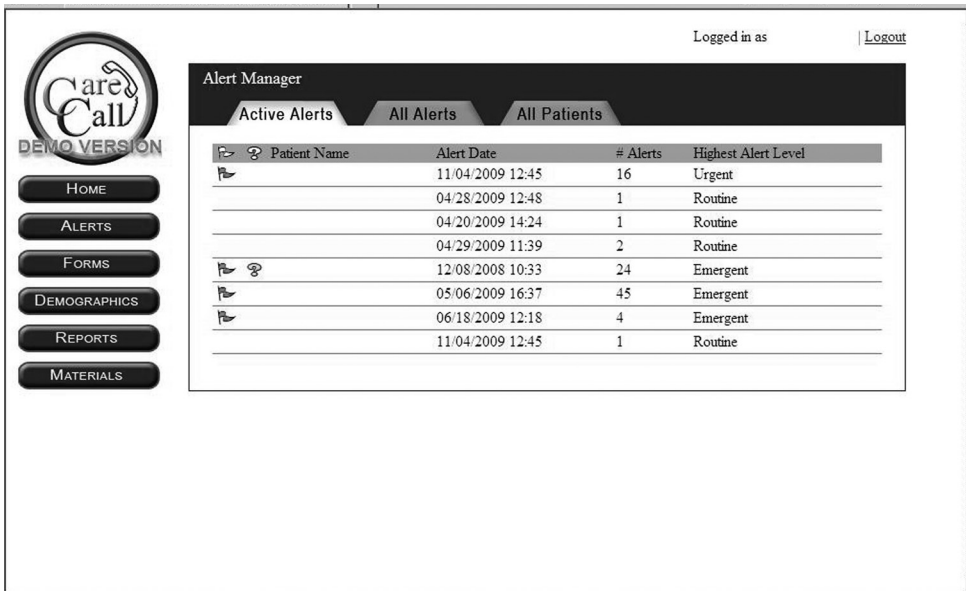


FIGURE 1 *Screenshot of the demo version of the Alert Manager, a Web-based tool for the nurse telerehabilitation coordinator, showing all active alerts (using mock data; mock names have been deleted).*

tracking form to document the length, content, and outcome of each contact with a user. We developed three levels of alerts for problems identified in Care Call: emergent (e.g., new skin problem), urgent (e.g., equipment problem), and routine (e.g., old problem that user wants to discuss with NTC). When an emergent medical problem is identified, Care Call directs users to contact their physician immediately or go to the emergency department or call 911. For urgent matters, Care Call tells users to see their physician as soon as possible and to ask for proper contact information for urgent medical problems at their next office visit (Figs. 1 and 2).

One of the major companions to Care Call is a Resource Book that all users receive at enrollment. Care Call and the NTC use this book in each encounter with participants. The Resource Book includes both local resources and informational resources for topics like medical supplies, mental and physical health providers, and personal care assistants. Included within the Resource Book is a set of user-friendly forms created by the research

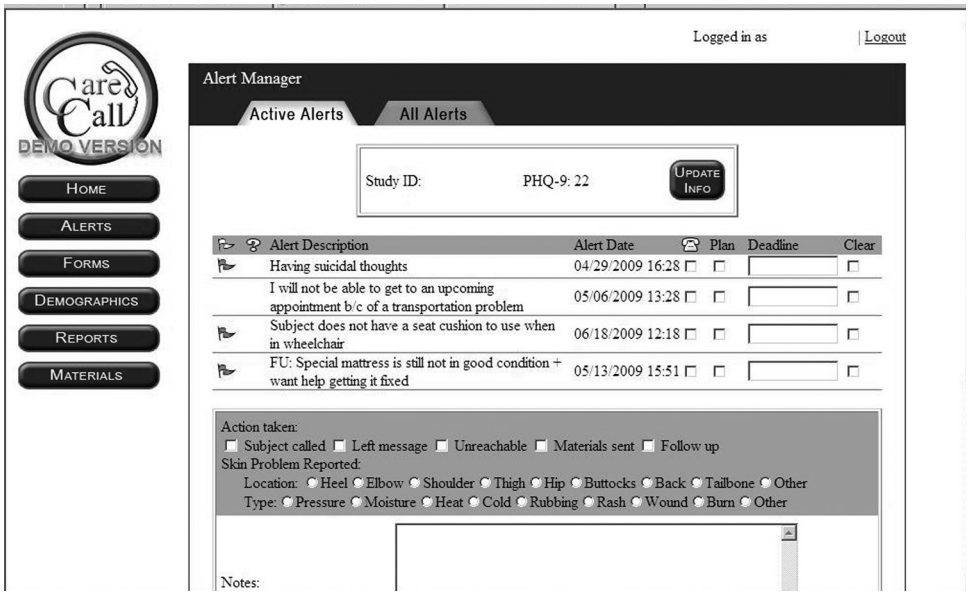


FIGURE 2 *Screenshot of active alerts for an individual mock subject in the demo version of the Alert Manager.*

team which Care Call reviews to help a user prepare for upcoming office visits to a health provider.

A Care Call Encounter

A Care Call user typically engages with Care Call weekly by receiving a call from TLC at an appointed time. Alternatively, a user may call into the system at any time to do their regular weekly call, report a new problem, hear the relaxation exercise, or leave a message for the NTC or technical staff. A typical conversation lasts 5 to 20 mins, depending on the patient's condition. After a unique password is entered, Care Call greets a user and a predetermined sequence of modules is delivered, as illustrated in Figure 3.

After an initial inquiry into any new skin problems, users hear different contents and questions in each call. User responses from previous encounters and the current encounter shape ensuing questions and feedback, such that content varies for each user and within each encounter. Throughout the entire conversation, Care Call alerts the NTC as needed and offers relevant patient audio clips, as well as tips from Care Call.

Pilot Testing

We used three types of testing in Care Call's development, the results of which have shaped both the final intervention and the clinical trial research protocol. The first was voice quality testing, which evaluates Care Call's ability to process various voice characteristics specific to our target population. Participants in the voice quality testing included five people with notable voice characteristics (ventilator use, pseudobulbar affect, dysarthria, soft-spoken speech, and stuffy nose) and one person with no notable voice characteristics. Testing the feasibility of Care Call for differing voice qualities was crucial for informing the criteria for participation in the clinical trial.

The second approach was pilot testing a beta version of the Care Call system with nine people with SCD to garner consumer feedback on content and to uncover any logic errors or technical difficulties. The Care Call team then revised the beta version of the Care Call system to create a final version for the trial. Subjects participated in six calls focusing on different scripts of the Care Call system and provided feedback during an in-depth interview.

All pilot testing procedures followed the approved protocol of the Boston University Medical Campus Institutional Review Board, conducted in accordance with the Declaration of the World Medical Association. Informed consent was obtained from all pilot participants.

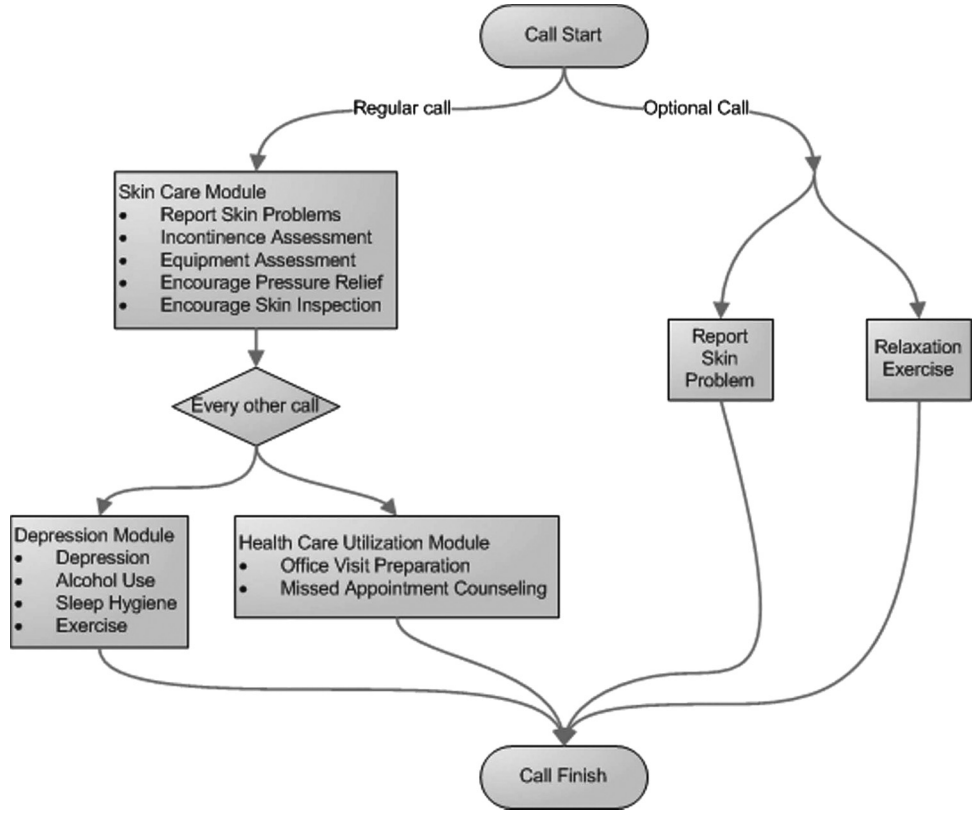


FIGURE 3 *Flow diagram depicting the architecture of Care Call.*

The final stage involved quality control testing to evaluate the final version of the Care Call intervention being deployed for the randomized controlled clinical trial, primarily for wording changes or technical errors. Three members of the development team assumed the identities of mock subjects with differing baseline characteristics. A report of unexplored areas of the system was distributed several times as the testing continued to guide testers in which untested pathways to follow with their responses. Pilot testing allowed the development team to address any possible issues before starting the trial.

RESULTS

The results of the voice quality testing showed that Care Call was able to understand all voice characteristics except very soft-spoken speech. In particular, the TLC system had difficulty hearing and understanding the responses of a soft-spoken user. After inquiring twice without being able to understand one of the response options provided, the system stated that there were technical difficulties and discontinued the call, per protocol. Voice quality results confirmed that the study could enroll subjects with a wide range of voice quality.

In the beta test, pilot study subjects found Care Call to be an acceptable intervention overall. They felt that Care Call could be useful particularly for people who were depressed, those with a new injury (SCI), and those without access to quality care. Several subjects mentioned repetition of information but recognized at the same time how this could be helpful for those with limited education in, for example, the topic of skin breakdown. Because of their feedback, we staggered content from week to week to shorten call times and modified content to acknowledge individual differences in perceived need for various recommendations. For instance, one pilot study subject felt that heel protectors did not apply to him/her and thus did not like that Care Call would continue to suggest use of them and inquire about obtaining them. Care Call was modified to include shoes as heel protectors while in the wheelchair and to have the nurse follow up with a user rather than having Care Call repeatedly promote the use of heel protectors.

Lastly, our quality control testing uncovered some logic problems in the scripts and identified content sequences that needed modification to maintain the utmost sensitivity to the varying circumstances of individuals in the clinical trial. For instance, a member of our research team first reported having two new skin problems and was then asked about the location of each. She imagined one on either side of her buttocks, yet Care Call did not specifically probe for this. When asking more about each skin problem, Care Call referred to each as being on the buttocks without differentiation between the right or the left buttock cheek. This line of questioning was confusing to the user, who did not know which problem to report on first.

Clinical experience suggested that most likely, skin problems on the same area of the body would be similar in cause and nature. Barring a total rewrite of this script and its logic, this issue was resolved most efficiently by adding a question with some instruction to users as follows: "Are you having a skin problem in just one area or in more than one area? If you're having problems on both sides of your body, for example both knees, count that as one area." Care Call was also revised to subsequently ask users if they had more than one skin problem in a particular area (in this case, the buttocks) to report on the worst skin problem, thus relieving any confusion or need for differentiation.

In terms of modifying content for sensitivity to individual differences, there was consensus that language needed to be softened related to feedback and education on adherence to skin care. Test callers felt that the weekly repetition of this content was reinforcement enough, such that the language itself should be less forceful, asking users if they would be willing to try to do more in the week ahead, rather than stating that they should. This change also allowed Care Call to be sensitive to the potential individual circumstances that might prevent a user from proper skin care, such as illness or a death in the family.

DISCUSSION

The Care Call telerehabilitation approach described in this article uses a ubiquitous instrument (the telephone) and, if shown to be efficacious, has the potential for widespread dissemination at low cost. Although automated, TLC programs can successfully emulate the educational and behavioral content, support, and conversational style of a human professional.[20,21,40] Research to date has shown that TLC technology in other disease areas has been able to successfully change health behavior and disease outcomes.[31] The preliminary results of our pilot testing are encouraging. The results of the voice quality testing showed that Care Call was able to understand all voice characteristics except very soft-spoken speech. This problem could, in some

cases, be alleviated with coaching via three-way calling on required voice volume levels. Ventilator use was not a problem, and the TLC system was able to appropriately confirm responses for the user with a stuffy nose and cough. And, importantly, pilot study subjects felt that Care Call could be useful particularly for people who are depressed, those with a new SCI, and those without access to quality care.

The results of an RCT evaluating the Care Call intervention will be available in 2011. The final intervention protocol was shaped in important ways by pilot testing the results. In the trial, intervention subjects receive weekly calls for 6 mos. Because of pilot testing, we were able to decrease the length of calls by removing some marginal content and staggering modules across calls. For instance, we removed most of a script that attempted to assess possible bowel leakage problems in detail because we found that this problem is so individualized that it is better to ask generally about whether the problem exists and then let the NTC follow up to further assess and make appropriate recommendations and referrals. We also ended up staggering modules to decrease call length from 30 mins to 15 mins, on average, with some individual variation based on choosing to hear optional content and/or need for follow-up from previously identified issues.

Different types of follow-up care have been developed to fill the gap in the continuum of rehabilitative care, especially for people with lifelong health needs, such as people with SCD. A systematic review of the literature in 2005 on follow-up care for people with SCI in the community showed that the most important methods have been telemedicine, outpatient consulting hours, home visits, or a combination of methods.[16] Because the quality of these studies was generally low, the authors were not able to draw conclusions as to the effect of these follow-up interventions on secondary conditions or long-term costs of care. Another systematic review of the literature in 2006 on preventing pressure ulcers showed that there were few well-designed RCTs following standardized criteria and providing data on cost-effectiveness,[41] suggesting that more need to be developed. As technology continues to advance, experts and clinicians are looking to the field of telehealth to fill this gap.[14] These technologies offer accessibility to potentially highly effective behavior change interventions at low cost. For instance, the TLC-Hypertension trial was the first telehealth chronic disease application to be evaluated in a randomized clinical trial, involving 267 elderly, poorly controlled hypertensive patients. Mean adjusted diastolic blood pressure decreased significantly; 69% of TLC users rated TLC in the upper quartile of satisfaction on a visual analog scale, whereas the cost per patient user for 6 mos of use was $32.50.[29]

There are limitations to TLC technology and the Care Call intervention that must be noted. First, because TLC is an automated system, technical errors of varying kinds can occur (e.g., the system does not always recognize the user's responses or accept responses that it is not prepared to hear). To address this, technical staff needs to be available to assist users, and all calls should be logged line by line on a server to record any errors in detail for proper follow-up. In addition, the system is limited to the conversational pathways that were designed; it cannot truly take each individual user's situation into account or discuss any topic that the user is interested in pursuing. This means that Care Call may be advocating for a certain behavior that is not appropriate or not perceived to be relevant to a particular user. The system includes statements to explain these limitations, but users may still ultimately be frustrated at times. The extent to which this limits the impact of the system is an empirical question about which we will learn from our clinical trial.

CONCLUSION

Care Call is designed to reduce the incidence and severity of secondary conditions such as pressure ulcers and depression, to be relevant across multiple consumer settings, and to facilitate long-term monitoring. The system is a low-risk, inexpensive intervention that could be widely disseminated. Results of pilot testing of Care Call demonstrated the intervention's feasibility for a wide variety of voice qualities and its general acceptability by consumers, with some minor content and logic changes. Pilot testing also allowed the team to uncover important content and logic changes that improved the system. These findings resulted in a stronger intervention and protocol for the clinical trial of Care Call.

Experience from the RCT will establish if people with SCD will use the Care Call system and if the intervention will successfully promote self-management in a cost-effective manner.

ACKNOWLEDGMENTS

We acknowledge Dr Dave Rosenblum and Mercedes Martinez for their contribution to the design and development of the Care Call intervention; Sippy Olarsch and Kelsey Williams for assisting with the development of the Web-based tools;

and the Medical Systems Information Unit team for technical guidance and support throughout.

REFERENCES

1. Beatty PW, Hagglund KJ, Neri MT, et al: Access to health care services among people with chronic or disabling conditions: patterns and predictors. *Arch Phys Med Rehabil* 2003;84:1417–25
2. Kemp B: Appendix N: depression as a secondary condition in people with disabilities, in Field M, Jette A (eds): *The Future of Disability in America*. Washington, DC: National Academies Press; 2007, pp 234–50
3. *Status of Persons with Spinal Cord Injury/Disease: Highlights*. Jackson Heights, NY, United Spinal Association, 2005
4. Garber S, Rintala D, Hart K, et al: Pressure ulcer risk in spinal cord injury: predictors of ulcer status over 3 years. *Arch Phys Med Rehabil* 2000;81:465–71
5. Chapman TM: *An Investigation of Risk Factors for Hidden Alcohol Abuse Among Adults with Physical Disabilities*. United States: ProQuest Information & Learning; 1998. Available at http://web.ebscohost.com/ehost/detail?hid=111&sid=671d1c23-bab0-477e-abc4-98e94b3a29ce%40sessionmgr110&vid=1&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#db=psyh&AN=1998-95018-194
6. Krause JS, Kemp B, Coker J, et al: Depression after spinal cord injury: relation to gender, ethnicity, aging, and socioeconomic indicators. *Arch Phys Med Rehabil* 2000;81:1099–109
7. Keri MT, Kroll T: Understanding the consequences of access barriers to health care: experiences of adults with disabilities. *Disabil Rehabil* 2003;25:85–96
8. Beatty PW, Bingham SC: Health care access and secondary conditions among people with spinal cord injury. In: *Proceedings of the 130th Meeting of APHA*; Philadelphia, PA, 2002. Available at http://apha.confex.com/apha/130am/techprogram/paper_48816.htm
9. Regan M, Teasell RW, Keast D, et al: Pressure ulcers following spinal cord injury, in Eng JJ, Teasell RW, Miller WC, et al (eds): *Spinal Cord Injury Rehabilitation Evidence. Version 3.0*, 2010. Available at www.scireproject.com
10. Orenczuk S, Slivinski J, Mehta S, et al: Depression following spinal cord injury, in Eng JJ, Teasell RW, Miller WC, et al (eds): *Spinal Cord Injury Rehabilitation Evidence. Version 3.0*, 2010. Available at www.scireproject.com
11. Iezzoni LI, Davis RB, Soukup J, et al: Satisfaction with quality and access to health care among people with disabling conditions. *Int J Qual Health Care* 2002;14:369–81
12. Canupp KC, Waites KB, DeVivo MJ, et al: Predicting compliance with annual follow-up evaluations in persons with spinal cord injury. *Spinal Cord* 1997; 35:314–9
13. Pinto BM, Friedman R, Marcus BH, et al: Effects of a computer-based, telephone-counseling system on physical activity. *Am J Prev Med* 2002;23:113–20
14. Ramelson HZ, Friedman RH, Ockene JK, et al: An automated telephone-based smoking cessation education and counseling system. *Patient Educ Couns* 1999;36:131–44
15. Healthy People 2010 (2nd ed.): *Understanding and Improving Health and Objectives for Improving Health (2 vols.)*. Washington, DC: US Government Printing Office; 2000. Available at http://www.healthypeople.gov/2010/Document/tableofcontents.htm#volume1
16. Bloemen-Vrencken JH, de Witte LP, Post MW, et al: Follow-up care for persons with spinal cord injury living in the community: a systematic review of interventions and their evaluation. *Spinal Cord* 2005;43:462–75
17. Lowery J: Telemedicine systems may help SCI patients get specialized care sooner. *Forum* 2004. Available at http://www.hsrd.research.va.gov/publications/internal/forum6_04.pdf
18. Dallolio L, Menarini M, China S, et al: THRIVE Project: Functional and clinical outcomes of telemedicine in patients with spinal cord injury. *Arch Phys Med Rehabil* 2008;89:2332–41
19. Regan M, Teasell RW, Keast D, et al: Pressure ulcers following spinal cord injury, in Eng JJ, Teasell RW, Miller WC, et al (eds): Spinal Cord Injury Rehabilitation Evidence. *Vancouver,* 2006; p 20.1–20.6. Available at http://www.icord.org/scire
20. Friedman RH, Stollerman JE, Mahoney DM, et al: The virtual visit: using telecommunications technology to take care of patients. *J Am Med Inform Assoc* 1997; 4:413–25
21. Migneault JP, Farzanfar R, Wright JA, et al: How to write health dialog for a talking computer. *J Biomed Inform* 2006;39:468–81
22. Greist JH: The computer as clinician assistant: assessment made simple. *Psychiatr Serv* 1998;49:467–8
23. Friedman RH, Stollerman J, Rozenblyum L, et al: A telecommunications system to manage patients with chronic disease. *Stud Health Technol Inform* 1998; 52:1330–4
24. Mahoney D, Tennstedt S, Friedman R, et al: An automated telephone system for monitoring the functional status of community-residing elders. *Gerontologist* 1999;39:229–34
25. Rubin A, Migneault JP, Marks L, et al: Automated telephone screening for problem drinking. *J Stud Alcohol* 2006;67:454–7
26. Farzanfar R: Computer telephony to improve adherence to antidepressants and clinical visits. *Ann Behav Med* 2003; Annual Meeting Supplement:S161
27. Adams WG, Fuhlbrigge AL, Miller CW, et al: TLC-asthma: an integrated information system for patient-centered monitoring, case management, and point-of-care decision support. *AMIA Annu Symp Proc* 2003;1–5
28. Friedman RH: Automated telephone conversations to assess health behavior and deliver behavioral interventions. *J Med Syst* 1998;22:95–102

29. Friedman RH, Kazis LE, Jette A, et al: A telecommunications system for monitoring and counseling patients with hypertension. Impact on medication adherence and blood pressure control. *Am J Hypertens* 1996;9:285–92

30. King A, Friedman RH, Marcus B, et al: Increasing regular physical activity via humans or automated technology: 12-month results of the CHAT trial. *Ann Behav Med* 2004;27:S044

31. Friedman R: *Telephone-Linked Communications (TLC) in Health Care: 20 Years Experience. AAAI Fall Symposium on Dialogue Systems in Health Communications*. Arlington, VA; 2004. Available at http://www.ccs.neu.edu/home/bickmore/dshc/friedmanslides.pdf

32. Delichatsios HK, Friedman RH, Glanz K, et al: Randomized trial of a "talking computer" to improve adults' eating habits. *Am J Health Promot* 2001;15:215–24

33. Prochaska JO, Velicer WF: The transtheoretical model of health behavior change. *Am J Health Promot* 1997;12:38–48

34. Bandura A: Health promotion by social cognitive means. *Health Educ Behav* 2004;31:143–64

35. Bandura A: *Social Learning Theory*. Englewood Cliffs, NJ: Prentice Hall; 1997

36. Sipski ML, Richards JS: Spinal cord injury rehabilitation: state of the science. *Am J Phys Med Rehabil* 2006;85:310–42

37. Jarvis KL, Friedman RH, Heeren T, et al: Older women and physical activity: using the telephone to walk. *Womens Health Issues* 1997;7:24–9

38. Hunkeler EM, Meresman JF, Hargreaves WA, et al: *Efficacy of nurse telehealth care and peer support in augmenting treatment of depression in primary care. Arch Fam Med* 2000;9:700–8

39. Wang PS, Simon GE, Avorn J, et al: *Telephone screening, outreach, and care management for depressed workers and impact on clinical and work productivity outcomes: a randomized controlled trial*. *JAMA* 2007;298:1401–11

40. Kaplan B, Farzanfar R, Friedman RH: Personal relationships with an intelligent interactive telephone health behavior advisor system: a multimethod study using surveys and ethnographic interviews. *Int J Med Inform* 2003;71:33–41

41. Reddy M, Gill SS, Rochon PA: Preventing pressure ulcers: a systematic review. *JAMA* 2006;296:974–84

AIDS Care
Vol. 23, No. 11, November 2011, 1483–1491

Operating characteristics of carbohydrate-deficient transferrin (CDT) for identifying unhealthy alcohol use in adults with HIV infection

Julia Ireland[a]*, Debbie M. Cheng[b,c], Jeffrey H. Samet[c,d], Carly Bridden[c], Emily Quinn[e,f] and Richard Saitz[c,g]

[a]*Department of Family Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA;* [b]*Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA;* [c]*Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA;* [d]*Department of Community Health Sciences, Boston University School of Public Health, Boston, MA, USA;* [e]*DM-STAT, Inc, Malden, MA, USA;* [f]*Data Coordinating Center, Boston University School of Public Health, Boston, MA, USA;* [g]*Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA*

(*Received 1 March 2010; final version received 16 February 2011*)

Unhealthy alcohol use (the spectrum of risky use through dependence) is common in HIV-infected persons, yet it can interfere with HIV medication adherence, may lower CD4 cell count, and can cause hepatic injury. Carbohydrate-deficient transferrin (CDT), often measured as %CDT, can detect heavy drinking but whether it does in people with HIV is not well established. We evaluated the operating characteristics of %CDT in HIV-infected adults using cross-sectional data from 300 HIV-infected adults with current or past alcohol problems. Past 30-day alcohol consumption was determined using the Timeline Followback (TLFB), a validated structured recall questionnaire, as the reference standard. Sensitivity and specificity of %CDT (at manufacturer's cut-off point of 2.6%) for detecting both "at-risk" ($\geq$4 drinks in a day or $>$7 drinks per week for women, $\geq$5 drinks in a day or $>$14 per week for men) and "heavy" drinking ($\geq$4 drinks in a day for women, $\geq$5 drinks in a day for men on at least seven days) were calculated. Receiver operating characteristic (ROC) curves were estimated to summarize the diagnostic ability of %CDT for distinguishing "at risk" and "heavy" levels of drinking. Exploratory analyses that stratified by gender and viral hepatitis infection were performed. Of 300 subjects, 103 reported current consumption at "at-risk" amounts, and 47 reported "heavy" amounts. For "at-risk" drinking, sensitivity of %CDT was 28% (95% confidence interval (CI) 19%, 37%), specificity 90% (95% CI 86%, 94%); area under the ROC curve (AUC) was 0.59. For "heavy" drinking, sensitivity was 36% (95% CI 22%, 50%), specificity 88% (95% CI 84%, 92%); AUC was 0.60. Sensitivity appeared lower among women and those with viral hepatitis; specificity was similar across subgroups. Among HIV-infected adults, %CDT testing yielded good specificity, but poor sensitivity for detecting "at-risk" and "heavy" alcohol consumption, limiting its clinical utility for detecting unhealthy alcohol use in this population.

Keywords: carbohydrate-deficient transferrin; CDT; alcohol; HIV

Background

HIV-infected populations in the USA have a high prevalence of unhealthy alcohol use, which can contribute to declines in their health (Cook et al., 2001; Galvan et al., 2002; Saitz, 2005; Samet, Phillips, Horton, Traphagen, & Freedberg, 2004). Alcohol can adversely affect immune function (Greiffenstein & Molina, 2008; Watzl & Watson, 1992), nutritional status (Lieber, 2003; Martin Villares et al., 2004) and adherence to medications (Braithwaite et al., 2005; Heckman, Catz, Heckman, Miller, & Kalichman, 2004). It can interfere with hepatic metabolism and is hepatotoxic. Furthermore, HIV co-infection with viral hepatitis is common (Shire et al., 2007; Tedaldi et al., 2003), and both HIV and viral hepatitis are adversely impacted by alcohol. For these reasons, detection of unhealthy alcohol use is important in the clinical care of HIV-infected individuals.

There is evidence that early intervention for unhealthy alcohol use can be effective (Kaner et al., 2009; Nilssen, 2004), but early clinical signs are often missed (Weisner & Matzger, 2003) and unhealthy alcohol use often goes undiagnosed by HIV health-care providers (Conigliaro et al., 2003). The National Institute on Alcohol Abuse and Alcoholism (NIAAA) advises screening for unhealthy alcohol use (National Institute on Alcohol Abuse and Alcoholism, 2007) by assessing for "at-risk" drinking amounts. These are drinking amounts that put individuals at increased risk of adverse health

*Corresponding author. Email: Julia.Ireland@fenwayhealth.org

ISSN 0954-0121 print/ISSN 1360-0451 online

http://dx.doi.org/10.1080/09540121.2011.565019
http://www.tandfonline.com

effects including alcohol use disorders (abuse and dependence).

Practice guidelines recommend screening by using questionnaires. Biomarkers are not standard for universal alcohol screening because they are more costly and less sensitive for unhealthy use than questionnaires (Coulton et al., 2006). But some have recommended their use. Biomarkers can be useful for detecting unhealthy use among those who deny it, helping to confirm unhealthy use, or to facilitate discussions with patients about alcohol by having a seemingly more objective medical test in hand (Miller & Anton, 2004). Biomarkers are sometimes used for monitoring of heavy alcohol use (Anton, Lieber, Tabakoff, & CDTect Study Group, 2002).

A number of biomarkers are affected by unhealthy alcohol consumption, e.g., lipids and liver enzymes, but none has both sufficient sensitivity and specificity for screening. The most common biomarkers currently used for detection of unhealthy alcohol use include the liver enzyme γ-glutamyltransferase (GGT), which is neither particularly sensitive nor specific for detecting unhealthy alcohol use (Conigrave et al., 2002; Schmidt et al., 1997; Schwan et al., 2004). Other biomarkers are red blood cell mean corpuscular volume (MCV), and a protein produced by the liver, carbohydrate-deficient transferrin (CDT), often measured as %CDT. Of these, only CDT has high specificity for unhealthy alcohol use (Schwan et al., 2004), does not remain elevated as long after an episode of unhealthy drinking as GGT (Schmidt et al., 1997), and retains specificity when liver disease is present (e.g., primary biliary cirrhosis) (Arndt, Meier, Nauck, & Gressner, 2006).

CDT's performance for detecting unhealthy alcohol use in people with HIV has not been well-established. HIV-infected populations have a high prevalence of abnormal liver enzymes. CDT might therefore be more sensitive and less specific for unhealthy alcohol consumption than it is in populations without HIV infection or with undetermined HIV status.

CDT operates in a dose-dependent manner as a biomarker for detecting alcohol consumption (Schellenberg et al., 2005). If CDT were able to detect "at-risk" drinking (the spectrum of "at-risk" use through dependence) in HIV-infected individuals, it could be very useful clinically (Conigliaro et al., 2003). Therefore, our primary objective was to evaluate the diagnostic accuracy of CDT to distinguish between "at-risk" and lesser than "at-risk" drinking in adults with HIV infection.

Methods

Subjects

We studied cross-sectional data collected prospectively for an observational cohort study, the HIV-Longitudinal Interrelationships of Viruses and Ethanol (HIV-LIVE) study (Samet et al., 2007). Subjects ($n = 400$) were recruited from 2001 to 2003 in the Boston area mainly from clinical settings (Samet et al., 2007). Mean age was 43, 75% were male, 33% white, 41% black, 19% Hispanic, approximately two-thirds were "heterosexual," one-third "gay or homosexual," and 56% had a lifetime history of injection drug use.

All subjects met the following eligibility criteria: (1) current or past alcohol problems ($\geq$two affirmative responses to the CAGE alcohol screening questionnaire or diagnosis of alcohol abuse or dependence based on clinical assessment by a physician-investigator); (2) HIV infection documented by an ELISA antibody test confirmed by western blot; (3) ability to speak English or Spanish; and (4) at least one contact person to assist with follow-up. Eligible subjects provided written informed consent before enrollment.

Exclusion criteria for the HIV-LIVE study were score of <21 on the Mini-Mental State Exam or trained-interviewer assessment of individual's inability to comprehend informed consent or answer interview questions. Additional exclusion criteria were lack of baseline hepatitis C RNA (viral load) data, missing responses to the TLFB questions, insufficient blood sample quantity, and inability to match blood sample to same-day questionnaire data.

Assessments

Laboratory tests and interviews by trained research associates were done at baseline and every six months for at least 24 months and up to 42 months ending in 2006. Blood samples were collected for storage annually. The Federal Government provided a certificate of confidentiality for added protection for the research data. The study was approved by the Institutional Review Boards at Boston Medical Center and Beth Israel Deaconess Hospital.

The reference standard used for determining past 30-day alcohol consumption was the TLFB, a lengthy validated structured recall questionnaire for assessing alcohol consumption over time in research studies (Carey, Carey, Maisto, & Henson, 2004; Sacks, Drake, Williams, Banks, & Herrell, 2003; Sobell & Sobell, 1995). We evaluated the ability of %CDT and GGT to detect three levels of unhealthy alcohol consumption: "at-risk," "heavy," and "frequent

heavy'' drinking as determined by the reference standard TLFB.

We used the NIAAA definition of ''at-risk'' drinking (National Institute on Alcohol Abuse and Alcoholism, 2007), but placed it in the context of a 30-day period: ≥ 4 drinks in a day in past 30 days or >7 drinks per week on average for women; ≥ 5 drinks in a day in past 30 days or >14 drinks per week on average for men. We defined ''heavy'' drinking as at least seven days of drinking at ≥ 4 drinks in a day for women and ≥ 5 drinks in a day for men in past 30 days, amounts at which physiologic and/or organ damage become more likely (Conigrave et al., 2002; White, Altmann, & Nanchahal, 2002). We defined ''frequent heavy'' drinking as at least seven consecutive days of ''heavy'' drinking in past 30 days. We used the NIAAA definition of one drink: 12 ounces beer, 5 ounces of wine or 1.5 ounces of 80 proof liquor (containing approximately 14 grams or 0.6 fluid ounces of pure alcohol) (National Institute on Alcohol Abuse and Alcoholism, 2007).

Data collection at study entry included the following: age, gender, current viral hepatitis B and C infection (hepatitis B infection determined by the presence of surface antigen; hepatitis C infection determined by the presence of hepatitis C RNA). Follow-up interviews were conducted every six months and generally replicated the baseline interview content.

Sample selection and testing

This study allowed for the testing of 300 blood specimens for %CDT. To ensure adequate sample sizes for the estimation of sensitivity and specificity of %CDT for distinguishing between ''at-risk'' drinking and lesser than ''at-risk'' drinking, we sought to have the sample consist of approximately equal numbers of ''at-risk'' and lesser than ''at-risk'' drinkers. Sensitivity is the proportion of people who meet criteria for unhealthy alcohol use and who have a positive test result. Specificity is the proportion of people who do not meet the criteria who have a negative test result. Thus, all 103 HIV-LIVE subjects reporting ''at-risk'' drinking were included in the sample of 300, including 47 who reported ''heavy'' drinking amounts, 22 of whom also reported ''frequent heavy'' drinking.

All 67 HIV-LIVE subjects who reported consumption that was lesser than ''at-risk'' but not abstinent (no drinking in past 30 days) were included in the study sample. All of these were included in an attempt to have similar numbers with lower risk drinking as with abstinence. To complete the sample of 300 unique subjects, 130 were randomly selected from the remaining subjects, all of whom who reported abstinence. Subjects with multiple samples had a single sample randomly selected.

Selected blood samples were sent for %CDT and GGT analysis to a single laboratory, the Clinical Neurobiological Laboratory at the Medical University of South Carolina. CDT was measured as %CDT using the most recent iteration of the Axis-Shield turbidimetric immunoassay, %CDT-TIA, using the manufacturer's recommended cut-off point of 2.6% to define a positive test. GGT positive cut-off points were defined as ≥ 30U/L for women and ≥ 40U/L for men.

Analysis

All analyses were performed using SAS version 9.1.3 (Cary, NC, USA). The primary analysis was estimation of the sensitivity and specificity of %CDT (cut-off point of 2.6%) for detecting past 30-day ''at-risk,'' ''heavy,'' and ''frequent heavy'' drinking. Ninety-five percent exact binomial confidence intervals were calculated for sensitivity and specificity. Receiver operating characteristic (ROC) curves, plots of sensitivity versus 1-specificity across the range of possible cut points for %CDT, were estimated to summarize the overall diagnostic accuracy of %CDT as well as to evaluate the optimal cut-off point for distinguishing subjects with ''at-risk,'' ''heavy,'' and ''frequent heavy'' levels of drinking. The point of perfect classification (i.e., 100% sensitivity and specificity) is in the upper left hand corner of the plot. We defined the optimal cut-off point as the single point on the curve that is closest to the top left corner of the graph (i.e., the point that maximizes the combination of sensitivity and specificity). The area under the ROC curve (AUC) is a summary measure of the performance of %CDT and represents the probability of ranking a randomly chosen ''at-risk'' drinker above a randomly chosen lesser than ''at-risk'' drinker. An AUC of 1.0 represents a perfect test and while a value of 0.5 represents a test that does not discriminate better than chance. Positive and negative predictive values (PPV, NPV, the probability of having [PPV] or not having [NPV] the condition given the test result) were estimated to further describe the markers.

Exploratory analyses, which stratified by gender and viral hepatitis (hepatitis B or C or both versus neither), were performed to assess potential differences in the accuracy of %CDT across subgroups. Secondary analyses were also conducted to evaluate the accuracy of %CDT for detecting past 14-day ''at-risk'' drinking and of the marker GGT as a test for detecting past 30-day ''at-risk,'' ''heavy,'' and ''frequent heavy'' drinking in an HIV-infected cohort.

We also evaluated the performance of a test discussed in the literature which combines CDT and GGT: $0.8 \times \ln(\text{GGT}) + 1.3 \times \ln(\%\text{CDT}) = \gamma\%\text{CDT}$ (Sillanaukee & Olsson, 2001).

Results

Of the 400 subjects in the cohort, 394 met our study criteria, yielding 1123 available serum samples, of which we tested 300 from unique subjects. Table 1 provides characteristics of study subjects: 34% and 16% drank "at-risk" and "heavy" amounts, respectively; mean age was 44 years; 77% were men; and median CD4 count was 389. More than half were currently taking antiretroviral medication. More than half had viral hepatitis, mainly hepatitis C. Median GGT levels were above normal.

The estimated sensitivity of %CDT for detecting past 30-day "at-risk," "heavy," and "frequent heavy" drinking amounts was 28%, 36% and 41%, respectively (Table 2). Corresponding estimates for specificity were 90%, 88% and 86%, respectively. In analyses that stratified by gender, estimates of sensitivity of %CDT appeared lower for women compared to men; however, the differences were not statistically significant; estimates of specificity were similar for the two groups (Table 3). Sensitivity of %CDT was lower for the subjects with viral hepatitis compared to those without, although the difference was not statistically significant. Specificity did not appear to differ by viral hepatitis status.

The ROC curves for %CDT in detecting 30-day levels of unhealthy alcohol use in the overall sample appear in Figures 1–3. The ROC curves show that with increasing amounts of drinking, %CDT has greater diagnostic accuracy, i.e., the estimated AUC for %CDT was 0.59, 0.60, and 0.68 for 30-day "at-risk," "heavy," and "frequent heavy" drinking, respectively. The optimal %CDT cut-off points for detecting past 30-day "at-risk," "heavy," and "frequent heavy" drinking were 2.2 (sensitivity = 39%, specificity = 81%), 2.3 (sensitivity = 47%, specificity = 78%), and 2.0 (sensitivity = 73%, specificity = 62%), respectively.

Secondary analyses indicated that the sensitivity of %CDT (at the manufacturer's recommended cut-off point) for detecting past 14-day unhealthy drinking ranged from 21% to 39%, and specificity ranged from 86% to 94%. The sensitivity of GGT for detecting past 30-day unhealthy drinking ranged from 79% to 91% for detecting "at-risk," "heavy," and "frequent heavy" drinking amounts while the specificity ranged from 28% to 32%.

The positive and negative predictive values of %CDT were 60% and 70%, respectively, for detecting 30-day "at-risk" drinking and 58% and 74%, respectively, for detecting 14-day "at-risk" drinking. The positive and negative predictive values for detecting 30-day "heavy" drinking were 35% and 88%, respectively. The area under the ROC curve for the plot of γ%CDT sensitivity versus 1-specificity was 0.62 for detecting "at-risk" drinking, 0.63 for "heavy" drinking, and 0.73 for "frequent heavy" drinking.

Discussion

Among HIV-infected adults with alcohol problems, %CDT had poor overall accuracy for detecting unhealthy drinking. At the manufacturer's recommended cut-off point, %CDT had good specificity; however, the sensitivity was too low for %CDT to be a clinically useful screening test. At the optimal cut-off for %CDT in this sample, the diagnostic accuracy of %CDT did not improve sufficiently to make it a clinically useful screening test.

Although the sensitivity of %CDT for detecting "at-risk" drinking appeared higher for men and those without hepatitis, 31% and 43%, respectively, it remained unacceptably low for these subgroups. However, even in a population such as the HIV-LIVE cohort, with a high prevalence of unhealthy alcohol use, the PPV was only 60%. NPV of 70% was similarly not very high, consistent with the low observed sensitivity.

The sensitivity and specificity for detecting past 14-day "at-risk" drinking was similar to the 30-day data, meaning that %CDT did not perform any better even when drinking was more recent. The high specificity of %CDT could make it useful in sequential testing, i.e., as a follow-up test to a positive finding on a highly sensitive questionnaire that screened for unhealthy alcohol use. Such an approach would require further study of both the precision and cost-effectiveness of sequential testing for determining unhealthy alcohol consumption by patients with HIV.

In spite of its recognized shortcomings, we analyzed the operating characteristics of GGT for detecting unhealthy drinking in an HIV-infected cohort because it is a test whose results are commonly available in clinical practice. The sensitivity of GGT was high in our study, much higher than what has been generally reported in the literature where GGT has been shown to have similar or even lower sensitivity than %CDT in samples of undetermined HIV status. It is common for GGT to be elevated in HIV-infected individuals. The high median GGT in

Table 1. Characteristics of HIV-infected subjects with current or past alcohol problems ($n = 300$).

		Past 30-day alcohol consumption			
				Subsets of "at-risk"	
Characteristic	Overall	lesser than "at-risk"	"at-risk"[a]	"heavy"[b]	"frequent heavy"[c] subset of "heavy"
n (% of total)	300 (100%)	197 (66%)	103 (34%)	47 (16%)	22 (7%)
Age (years) mean (SD)	44 (8)	45 (8)	42 (7)	40 (6)	40 (6)
	n (%) of column		*n* (%) of row		
Women	68 (23%)	45 (66%)	23 (34%)	15 (22%)	5 (7%)
Men	232 (77%)	152 (66%)	80 (34%)	32 (14%)	17 (7%)
Current hepatitis[d] B or C	168 (56%)	105 (63%)	63 (38%)	33 (20%)	15 (9%)
No hepatitis C	147 (49%)	100 (68%)	47 (32%)	18 (12%)	8 (5%)
Current hepatitis C	153 (51%)	97 (63%)	56 (37%)	29 (19%)	14 (9%)
Taking anti-retroviral medication	181 (60%)	133 (73%)	48 (27%)	16 (9%)	9 (5%)
			Median[e] (Q1–Q3)[f] of row		
CD4 cells/mm^3	389 (231–582)	410 (250–608)	321 (187–543)	289 (160–474)	214 (95–387)
% CDT	1.8 (1.5–2.3)	1.8 (1.5–2.1)	2 (1.6–2.7)	2.1 (1.5–2.9)	2.3 (1.7–2.9)
GGT	64 (35–169)	58 (30–146)	105 (43–202)	126 (42–213)	149 (77–341)

[a]"at-risk" drinking uses N.I.A.A.A. definition of ≥ 4 drinks in a day or > 7 drinks per week on average for women; ≥ 5 drinks in a day or > 14 drinks per week on average for men, but placed in past 30 days.
[b]"heavy" drinking: ≥ 4 drinks in a day for women; ≥ 5 drinks in a day for men on at least 7 days in past 30 days.
[c]"frequent" heavy: at least seven consecutive days of drinking "heavy" amounts in past 30 days.
[d]Current hepatitis determined from $n = 299$. Hepatitis B infection determined by the presence of surface antigen; hepatitis C infection determined by the presence of hepatitis C RNA, i.e., positive hepatitis C viral load. Seven subjects had no record of hepatitis B surface antigen and one of those had undetectable hepatitis C viral load.
[e]CD4 median determined from $n = 287$. Thirteen subjects had no record of same-day CD4 count.
[f]Q1 = 25th percentile, Q3 = 75th percentile.

Table 2. Sensitivity and specificity of %CDT and of GGT for detecting 30-days "at-risk," "heavy" and "frequent heavy" drinking.

			Subsets of "at-risk"			
	"at-risk"[a]		"heavy" (subset of "at-risk")[b]		"frequent heavy" (subset of "heavy")[c]	
Categories of unhealthy alcohol consumption	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Sample size for estimate (% of total sample)	103 (34%)	197 (66%)	47 (16%)	253 (84%)	22 (7%)	278 (93%)
%CDT[d]	28% (19%, 37%)	90% (86%, 94%)	36% (22%, 50%)	88% (84%, 92%)	41% (20%, 61%)	86% (82%, 90%)
GGT[e]	82% (74%, 9%)	32% (26%, 9%)	79% (67%, 0%)	28% (23%, 4%)	91% (79%, 100%)	29% (24%, 34%)

[a]"at-risk" drinking: $\geq$4 drinks in a day or $>$7 drinks per week on average for women; $\geq$5 drinks in a day or $>$14 drinks per week on average for men.
[b]"heavy" drinking: $\geq$4 drinks in a day for women; $\geq$5 drinks in a day for men on at least 7 days in past 30 days.
[c]"frequent heavy" drinking: at least seven consecutive days of drinking "heavy" amounts.
[d]%CDT positive cut-off point was $\geq$2.6%
[e]GGT positive cut-off points were $\geq$30 U/L for women and $\geq$40 U/L for men

Table 3. Sensitivity and specificity of %CDT stratified by gender and viral hepatitis.

			Subsets of "at-risk"			
	"at-risk"[a]		"heavy"(subset of "at-risk")[b]		"frequent heavy" (subset of "heavy")[c]	
	Sensitivity (95% CI[d])	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Women $n=68$	0.17 (0.02, 0.33)	0.91 (0.83, 0.99)	0.20 (0.00, 0.40)	0.91 (0.83, 0.98)	0.20 (0.00, 0.55)	0.89 (0.81, 0.97)
Men $n=232$	0.31 (0.21, 0.41)	0.90 (0.85, 0.95)	0.44 (0.27, 0.61)	0.87 (0.82, 0.92)	0.47 (0.23, 0.71)	0.85 (0.80, 0.90)
No hepatitis $n=131$	0.43 (0.27, 0.58)	0.91 (0.85, 0.97)	0.64 (0.39, 0.89)	0.86 (0.80, 0.93)	0.57 (0.21, 0.94)	0.83 (0.76, 0.90)
Hepatitis B or C $n=168$	0.19 (0.09, 0.29)	0.90 (0.84, 0.95)	0.24 (0.10, 0.39)	0.89 (0.84, 0.94)	0.33 (0.09, 0.57)	0.88 (0.83, 0.93)

[a]"at-risk" drinking: $\geq$4 drinks in a day or $>$7 drinks per week on average for women; $\geq$5 drinks in a day or $>$14 drinks per week on average for men.
[b]"heavy" drinking: $\geq$4 drinks in a day for women; $\geq$5 drinks in a day for men on at least 7 days in past 30 days.
[c] "frequent heavy" drinking: at least seven consecutive days of drinking "heavy" amounts.
[d]CI = confidence interval.
Notes: Hepatitis B = hepatitis B surface antigen; positive Hepatitis C = presence of detectable viral load. Hepatitis strata calculated from $n=299$ (Seven subjects had no record of hepatitis B surface antigen and one of those had undetectable hepatitis C viral load.)

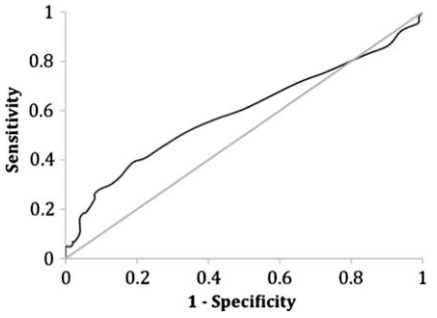


Figure 1. %CDT sensitivity vs. 1-specificity for detecting 30-day "at-risk" drinking.
Notes: Area Under the Curve (AUC): 0.59. The optimal%CDT cut-off point to discriminate 30-day at-risk drinking is 2.2. Sensitivity = 39%, Specificity = 81%.

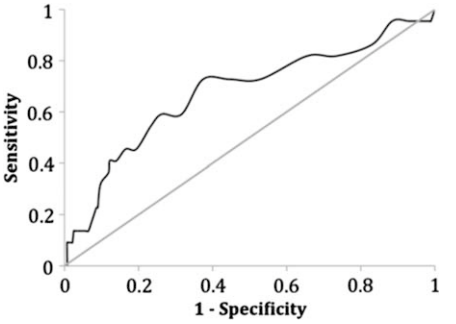


Figure 3. %CDT sensitivity vs. 1-specificity for detecting 30-day "frequent heavy" drinking.
Notes: Area Under the Curve (AUC): 0.68. The optimal%CDT cut-off point to discriminate 30-day frequent heavy drinking is 2.0. Sensitivity = 73%, Specificity = 62%.

our sample, likely of multifactorial etiology, resulted in a very high sensitivity of GGT for detecting unhealthy alcohol use in our sample. This high sensitivity of GGT came at a cost of specificity. Specificity was too low for GGT to be useful as a screening test for unhealthy alcohol use. Also, we found that γ %CDT does not provide a clinically significant advantage over %CDT alone in HIV-infected individuals.

The operating characteristics of %CDT for detecting the spectrum of unhealthy alcohol use by HIV-infected adults has not received adequate attention. Many studies with adults with undetermined HIV status have found %CDT to have greater sensitivity than observed in this study (Bortolotti, De Paoli, & Tagliaro, 2006). These studies included subjects who were drinking more heavily and tested whether %CDT could detect these heavier amounts (Anttila, Jarvi, Latvala, & Niemela, 2004; Chrostek, Cylwik, Szmitkowski, & Korcz, 2006; Zierau et al., 2005). When studies that evaluated %CDT for detecting daily heavy alcohol consumption are excluded, as was done in a systematic review by Koch et al (2004), the range of %CDT sensitivity for detecting lower levels of consumption is similar to our findings.

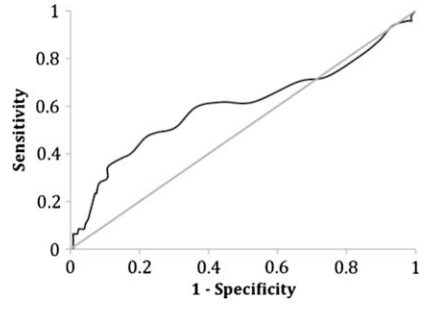


Figure 2. %CDT sensitivity vs. 1-specificity for detecting 30-day "heavy" drinking.
Notes: Area Under the Curve (AUC): 0.60. The optimal%CDT cut-off point to discriminate 30-day heavy drinking is 2.3. Sensitivity = 47%, Specificity = 78%.

This study had its limitations. First, it included a small number of people with "frequent heavy" drinking. On the other hand, this may also be viewed as a strength since we studied an adequate number of people with the condition that is the target of greater importance for screening – the spectrum of unhealthy alcohol use – thereby avoiding the spectrum bias that limited a number of prior studies. Nevertheless the sample size limited the precision of our estimates, particularly in subgroup analyses (e.g., by gender and viral hepatitis). Second, we did not study subjects who had never had alcohol problems. However, we did include subjects with abstinence and those who consumed amounts of alcohol in a range from moderate to heavy use. Third, since we were studying a biomarker, one might question what the reference standard should be. The common biomarkers mentioned, GGT and MCV, lack specificity, and self report may be biased. We chose a widely agreed-upon and validated self-report research tool; our interviews were conducted with assurance of confidentiality and coincide with both an alcohol breath test and blood testing to encourage truth telling.

This study had a number of strengths. We did not exclude people who used ART medications, had liver disease and other co-morbidities. As such, our sample provided a generalizable test of %CDT in HIV-infected adults. In addition, our sample was relatively young, CD4 counts were relatively high, HIV RNA viral loads were relatively low, and so the performances of %CDT and GGT in our study are not likely attributable to very advanced HIV disease.

These findings suggest that %CDT is not sufficiently sensitive for use in screening for unhealthy

alcohol use by people with HIV infection. Next steps for research might include testing other biomarkers for this purpose. Numerous self-report questionnaires have been validated for detecting unhealthy alcohol use. These will likely remain the least expensive, most accurate and most easily implementable tools for screening patients with HIV for unhealthy alcohol use.

Acknowledgements

This study was funded by the following grants: T32HP10028-06, R01AA013216, K24AA015674, and General Clinical Research Centers at Boston University School of Medicine (M01 RR00533) and Beth Israel Deaconess Medical Center (M01 RR01032). We acknowledge Joseph Novotny, MPH, for his assistance in manuscript preparation.

References

Anton, R.F., Lieber, C., Tabakoff, B., & CDTect Study Group. (2002). Carbohydrate-deficient transferrin and gamma-glutamyltransferase for the detection and monitoring of alcohol use: Results from a multisite study. *Alcoholism, Clinical and Experimental Research, 26*(8), 1215–1222.

Anttila, P., Jarvi, K., Latvala, J., & Niemela, O. (2004). Method-dependent characteristics of carbohydrate-deficient transferrin measurements in the follow-up of alcoholics. *Alcohol and Alcoholism, 39*(1), 59–63.

Arndt, T., Meier, U., Nauck, M., & Gressner, A.M. (2006). Primary biliary cirrhosis is not a clinical condition for increased carbohydrate-deficient transferrin: Experience with four independent CDT analysis methods. *Clinica Chimica Acta, 372*(1–2), 184–187.

Bortolotti, F., De Paoli, G., & Tagliaro, F. (2006). Carbohydrate-deficient transferrin (CDT) as a marker of alcohol abuse: A critical review of the literature 2001–2005. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences, 841*(1–2), 96–109.

Braithwaite, R.S., McGinnis, K.A., Conigliaro, J., Maisto, S.A., Crystal, S., Day, N., … Justice, A.C. (2005). A temporal and dose-response association between alcohol consumption and medication adherence among veterans in care. *Alcoholism Clinical and Experimental Research, 29*(7), 1190–1197.

Carey, K.B., Carey, M.P., Maisto, S.A., & Henson, J.M. (2004). Temporal stability of the timeline followback interview for alcohol and drug use with psychiatric outpatients. *Journal of Studies on Alcohol, 65*(6), 774–781.

Chrostek, L., Cylwik, B., Szmitkowski, M., & Korcz, W. (2006). The diagnostic accuracy of carbohydrate-deficient transferrin, sialic acid and commonly used markers of alcohol abuse during abstinence. *Clinica Chimica Acta, 364*(1–2), 167–171.

Conigliaro, J., Gordon, A.J., McGinnis, K.A., Rabeneck, L., Justice, A.C., & Veterans Aging Cohort 3-Site Study. (2003). How harmful is hazardous alcohol use and abuse in HIV infection: Do health care providers know who is at risk? *Journal of Acquired Immune Deficiency Syndromes, 33*(4), 521–525.

Conigrave, K.M., Degenhardt, L.J., Whitfield, J.B., Saunders, J.B., Helander, A., Tabakoff, B., & WHO/ISBRA Study Group. (2002). CDT, GGT, and AST as markers of alcohol use: The WHO/ISBRA collaborative project. *Alcoholism, Clinical and Experimental Research, 26*(3), 332–339.

Cook, R.L., Sereika, S.M., Hunt, S.C., Woodward, W.C., Erlen, J.A., & Conigliaro, J. (2001). Problem drinking and medication adherence among persons with HIV infection. *Journal of General Internal Medicine, 16*(2), 83–88.

Coulton, S., Drummond, C., James, D., Godfrey, C., Bland, J.M., Parrott, S., … Stepwice Research Team. (2006). Opportunistic screening for alcohol use disorders in primary care: Comparative study. *BMJ, 332*(7540), 511–517.

Galvan, F.H., Bing, E.G., Fleishman, J.A., London, A.S., Caetano, R., Burnam, M.A., … Shapiro, M. (2002). The prevalence of alcohol consumption and heavy drinking among people with HIV in the United States: Results from the HIV cost and services utilization study. *Journal of Studies on Alcohol, 63*(2), 179–186.

Greiffenstein, P., & Molina, P.E. (2008). Alcohol-induced alterations on host defense after traumatic injury. *The Journal of Trauma, 64*(1), 230–240.

Heckman, B.D., Catz, S.L., Heckman, T.G., Miller, J.G., & Kalichman, S.C. (2004). Adherence to antiretroviral therapy in rural persons living with HIV disease in the United States. *AIDS Care, 16*(2), 219–230.

Kaner, E.F., Dickinson, H.O., Beyer, F., Pienaar, E., Schlesinger, C., Campbell, F., … Heather, N. (2009). The effectiveness of brief alcohol interventions in primary care settings: A systematic review. *Drug and Alcohol Review, 28*(3), 301–323.

Koch, H., Meerkerk, G.J., Zaat, J.O., Ham, M.F., Scholten, R.J., & Assendelft, W.J. (2004). Accuracy of carbohydrate-deficient transferrin in the detection of excessive alcohol consumption: A systematic review. *Alcohol and Alcoholism, 39*(2), 75–85.

Lieber, C.S. (2003). Relationships between nutrition, alcohol use, and liver disease. *Alcohol Research & Health, 27*(3), 220–231.

Martin Villares, C., Dominguez Calvo, J., San Roman Carbajo, J., Fernandez Pello, M.E., Pomar Blanco, P., & Tapia Risueno, M. (2004). Ingesta excesiva de alcohol, malnutricion y cancer de cabeza y cuello [Heavy alcohol intake, malnutrition and head and neck cancer patients]. *Nutricion Hospitalaria, 19*(6), 348–352.

Miller, P.M., & Anton, R.F. (2004). Biochemical alcohol screening in primary health care. *Addictive Behaviors, 29*(7),1427–1437.

National Institute on Alcohol Abuse and Alcoholism. (2007). *Helping patients who drink too much: A clinician's guide: Updated 2005 edition*. Bethesda, MD: National Institutes of Health.

Nilssen, O. (2004). Long-term effect of brief intervention in at-risk alcohol drinkers: A 9-year follow-up study. *Alcohol and Alcoholism, 39*(6), 548–551.

Sacks, J.A., Drake, R.E., Williams, V.F., Banks, S.M., & Herrell, J.M. (2003). Utility of the time-line follow-back to assess substance use among homeless adults. *The Journal of Nervous and Mental Disease, 191*(3), 145–153.

Saitz, R. (2005). Clinical practice. Unhealthy alcohol use. *The New England Journal of Medicine, 352*(6), 596–607.

Samet, J.H., Cheng, D.M., Libman, H., Nunes, D.P., Alperen, J.K., & Saitz, R. (2007). Alcohol consumption and HIV disease progression. *Journal of Acquired Immune Deficiency Syndromes, 46*(2), 194–199.

Samet, J.H., Phillips, S.J., Horton, N.J., Traphagen, E.T., & Freedberg, K.A. (2004). Detecting alcohol problems in HIV-infected patients: Use of the CAGE questionnaire. *AIDS Research and Human Retroviruses, 20*(2), 151–155.

Schellenberg, F., Schwan, R., Mennetrey, L., Loiseaux, M.N., Pages, J.C., & Reynaud, M. (2005). Dose-effect relation between daily ethanol intake in the range 0–70 grams and %CDT value: Validation of a cut-off value. *Alcohol and Alcoholism, 40*(6), 531–534.

Schmidt, L.G., Schmidt, K., Dufeu, P., Ohse, A., Rommelspacher, H., & Muller, C. (1997). Superiority of carbohydrate-deficient transferrin to gamma-glutamyltransferase in detecting relapse in alcoholism. *The American Journal of Psychiatry, 154*(1), 75–80.

Schwan, R., Loiseaux, M.N., Schellenberg, F., Albuisson, E., Favre, J.D., Rigaud, A., ... Reynaud, M. (2004). Multicenter validation study of the %CDT TIA kit in alcohol abuse and alcohol dependence. *Alcoholism, Clinical and Experimental Research, 28*(9), 1331–1337.

Shire, N.J., Rouster, S.D., Stanford, S.D., Blackard, J.T., Martin, C.M., Fichtenbaum, C.J., & Sherman, K.E. (2007). The prevalence and significance of occult hepatitis B virus in a prospective cohort of HIV-infected patients. *Journal of Acquired Immune Deficiency Syndromes, 44*(3), 309–314.

Sillanaukee, P., & Olsson, U. (2001). Improved diagnostic classification of alcohol abusers by combining carbohydrate-deficient transferrin and gamma-glutamyltransferase. *Clinical Chemistry, 47*(4), 681–685.

Sobell, L.C., & Sobell, M.B. (1995). *Alcohol timeline followback (TLFB) users' manual*. Toronto, Canada: Addiction Research Foundation.

Tedaldi, E.M., Hullsiek, K.H., Malvestutto, C.D., Arduino, R.C., Fisher, E.J., Gaglio, P.J., ... Terry Beirn Community Programs for Clinical Research on AIDS. (2003). Prevalence and characteristics of hepatitis C virus coinfection in a human immunodeficiency virus clinical trials group: The Terry Beirn community programs for clinical research on AIDS. *Clinical Infectious Diseases, 36*(10), 1313–1317.

Watzl, B., & Watson, R.R. (1992). Role of alcohol abuse in nutritional immunosuppression. *The Journal of Nutrition, 122*(Suppl. 3), 733–737.

Weisner, C., & Matzger, H. (2003). Missed opportunities in addressing drinking behavior in medical and mental health services. *Alcoholism, Clinical and Experimental Research, 27*(7), 1132–1141.

White, I.R., Altmann, D.R., & Nanchahal, K. (2002). Alcohol consumption and mortality: Modelling risks for men and women at different ages. *BMJ, 325*(7357), 191–197.

Zierau, F., Hardt, F., Henriksen, J.H., Holm, S.S., Jorring, S., Melsen, T., & Becker, U. (2005). Validation of a self-administered modified CAGE test (CAGE-C) in a somatic hospital ward: Comparison with biochemical markers. *Scandinavian Journal of Clinical and Laboratory Investigation, 65*(7), 615–622.

Original Article

Structural and Reliability Analysis of a Patient Satisfaction With Cancer-Related Care Measure

A Multisite Patient Navigation Research Program Study

Pascal Jean-Pierre, PhD, MPH[1,2]; Kevin Fiscella, MD, MPH[3]; Karen M. Freund, MD, MPH[4]; Jack Clark, PhD[5]; Julie Darnell, PhD, MHSA[6]; Alan Holden, PhD[7]; Douglas Post, PhD[8]; Steven R. Patierno, PhD[9]; Paul C. Winters, MS[3]; and the Patient Navigation Research Program Group

BACKGROUND: Patient satisfaction is an important outcome measure of quality of cancer care and 1 of the 4 core study outcomes of the National Cancer Institute (NCI)-sponsored Patient Navigation Research Program to reduce race/ethnicity-based disparities in cancer care. There is no existing patient satisfaction measure that spans the spectrum of cancer-related care. The objective of this study was to develop a Patient Satisfaction With Cancer Care measure that is relevant to patients receiving diagnostic/therapeutic cancer-related care. **METHODS:** The authors developed a conceptual framework, an operational definition of Patient Satisfaction With Cancer Care, and an item pool based on literature review, expert feedback, group discussion, and consensus. The 35-item Patient Satisfaction With Cancer Care measure was administered to 891 participants from the multisite NCI-sponsored Patient Navigation Research Program. Principal components analysis (PCA) was conducted for latent structure analysis. Internal consistency was assessed using Cronbach coefficient alpha (α). Divergent analysis was performed using correlation analyses between the Patient Satisfaction With Cancer Care, the Communication and Attitudinal Self-Efficacy–Cancer, and demographic variables. **RESULTS:** The PCA revealed a 1-dimensional measure with items forming a coherent set explaining 62% of the variance in patient satisfaction. Reliability assessment revealed high internal consistency (α ranging from 0.95 to 0.96). The Patient Satisfaction With Cancer Care demonstrated good face validity, convergent validity, and divergent validity, as indicated by moderate correlations with subscales of the Communication and Attitudinal Self-Efficacy–Cancer (all $P < .01$) and nonsignificant correlations with age, primary language, marital status, and scores on the Rapid Estimate of Adult Literacy in Medicine Long Form (all $P > .05$). **CONCLUSIONS:** The Patient Satisfaction With Cancer Care is a valid tool for assessing satisfaction with cancer-related care for this sample. ***Cancer* 2011;117:854–61.** *© 2010 American Cancer Society.*

KEYWORDS: cancer, disparities, satisfaction, psychometrics, measurement, patient navigation, race-ethnicity.

Patient satisfaction reflects a core dimension of healthcare quality and patient-centered care.[1-3] Patient satisfaction indicates the extent to which patients' healthcare experiences match their expectations.[4,5] The construct of patient satisfaction has been linked to health status, quality of life, adherence to recommended treatment and medical advice including cancer treatment, initiation of complaints, and patient-healthcare provider communication in the clinical dyad.[6-16] Patient satisfaction with care represents an important outcome measure for healthcare in general and cancer care in particular.[17]

Corresponding author: Pascal Jean-Pierre, PhD, MPH, Department of Pediatrics, University of Miami Miller School of Medicine, Department of Pediatrics and Sylvester Comprehensive Cancer Center, P. O. Box 016820 (D-820), Miami, FL 33101; PJean-Pierre@med.miami.edu

[1]Department of Pediatrics, University of Miami Miller School of Medicine, Miami, Florida; [2]Sylvester Comprehensive Cancer Center, University of Miami, Miami, Florida; [3]University of Rochester Medical Center, Rochester, New York; [4]Boston University School of Medicine, Boston, Massachusetts; [5]Edith Norse Rogers Memorial Veterans Hospital, Bedford, Massachusetts; [6]University of Illinois, Chicago School of Public Health, Chicago, Illinois; [7]University of Texas Health Science Center, San Antonio, Texas; [8]The Ohio State University Comprehensive Cancer Center, Columbus, Ohio; [9]George Washington University, Cancer Institute, Washington, District of Columbia

The Patient Navigation Research Group consists of Fred Snyder, PhD, NOVA Research; Emmanuel Taylor, PhD, Center to Reduce Cancer Health Disparities, National Cancer Institute; Richard G. Roetzheim, MD, MSPH, H. Lee Moffitt Cancer Center and Research Institute; Victoria Warren-Mears, PhD, RD, LD, Northwest Tribal Epidemiology Center; Desiree Rivers, PhD, MSPH, University of South Florida, H. Lee Moffitt Cancer Center; and Nancy LaVerda, MPH, George Washington University, Cancer Institute, Washington, DC.

DOI: 10.1002/cncr.25501, **Received:** March 10, 2010; **Revised:** May 3, 2010; **Accepted:** May 25, 2010, **Published online** October 4, 2010 in Wiley Online Library (wileyonlinelibrary.com)

Patient satisfaction is 1 of the primary study outcomes of the National Cancer Institute (NCI)-supported Patient Navigation Research Program to reduce disparities in cancer care for individuals from racial/ethnic minorities and lower socioeconomic groups. The Patient Navigation Research Program involves 9 independent research programs operating under cooperative agreements with the NCI Center to Reduce Cancer Health Disparities to evaluate the impact of patient navigation on outcomes among patients with cancer screening abnormalities or diagnosed cancer.[18]

Although there are numerous patient satisfaction measures, including several measures related to cancer treatment, none of these measures spans the spectrum of cancer-related care from screening to treatment of diagnosed cancer.[19-25] For example, the widely used EORTC-IN-PATSAT32 is designed to assess satisfaction with the inpatient cancer care, and FAMCARE assesses satisfaction among those with advanced cancer.[20,24]

In the present study, we aimed to develop a Patient Satisfaction With Cancer Care measure that had: 1) sufficient breadth (ie, addressing satisfaction with care during evaluation of screening abnormalities and treatment), 2) the ability to address many of the challenges confronted by poor and minority individuals receiving cancer-related care, and 3) relevance for evaluation of care among both navigated and un-navigated patients.

MATERIALS AND METHODS

Development of the Patient Satisfaction With Cancer Care

The scale development team included investigators from different Patient Navigation Research Program sites with content and technical expertise in clinical care of patients from diverse cultural and socioeconomic backgrounds, as well as measurement development and psychometrics. The team reviewed existing patient satisfaction measures, considered various domains of satisfaction (access/logistical, interpersonal/relational, communicational/informational, and coordination of care), and selected and modified existing items for inclusion in the new Patient Satisfaction With Cancer Care scale. One additional item was administered only to participants with a confirmed diagnosis of cancer: "My treatment was explained in a way I could understand."

Response Options and Scoring

Patients responded to each scale item on a 5-point Likert scale (1 = Strongly Agree to 5 = Strongly Disagree). A total scale score was obtained by adding scores on all items, with lower scores indicating higher satisfaction with cancer care.

Participants

The Patient Navigation Research Program methods have been previously published.[18] Briefly, the Patient Navigation Research Program is a cooperative program funded by the NCI and the American Cancer Society to rigorously evaluate the role and benefits of patient navigation among participants with abnormal cancer screening findings or diagnosed cancer-breast, cervical, colorectal, or prostate cancer within 9 largely racial/ethnic minority and low-income communities across the country. Study design and type of cancer differ by participating site.

The satisfaction items were administered to a subsample of the 8075 participants in the Patient Navigation Research Program. In all, 891 English-fluent participants from the multisite NCI-sponsored Patient Navigation Research Program completed the Patient Satisfaction With Cancer Care measure. Survey participants were similar in age, but more likely to be female, minority, lower income, and less educated.

Procedures

Medical staff at the Patient Navigation Research Program recruiting sites (eg, clinics or hospitals) was informed about the study and referred eligible patients to speak with a trained research assistant or patient navigator about participating in the study. To minimize possible effects of low literacy, surveys were read out loud to participants in English.

Eligibility and Exclusion Criteria

Eligibility for the present study included having an abnormal breast, cervical, colorectal, and prostate cancer test finding or a new diagnosis of these cancers without any prior history of cancer treatment other than nonmelanoma skin cancer.[18]

Additional Measures

Demographic characteristics

Demographic characteristics included age, sex, race, ethnicity, primary language, income, education, marital status, and whether the patient received care related to evaluation of cancer screening abnormalities or treatment of cancer, and type of cancer being evaluated or treated (breast, cervical, colorectal, or prostate).

Communication and Attitudinal Self-Efficacy–Cancer

The Communication and Attitudinal Self-Efficacy–Cancer is a psychometrically validated multidimensional measure (ie, understanding and participating in care, maintaining a positive attitude, seeking and obtaining information) of communication and attitude. Structural analysis of the Communication and Attitudinal Self-Efficacy–Cancer revealed high internal consistency and construct validity.[26] Given overlap in constructs, we expected that the Patient Satisfaction With Cancer Care would correlate with the Communication and Attitudinal Self-Efficacy–Cancer.

Data Analysis

Dimensionality analysis of the Patient Satisfaction With Cancer Care

Latent structural and psychometric validation analyses were conducted using the SPSS version 17.0 statistical software package for Microsoft Windows (SPSS Inc., Chicago, Ill). Data from our multisite sample were randomly divided into 2 separate datasets (Sample 1, $N_1 = 453$; Sample 2, $N_2 = 438$) using SPSS. One dataset was used to test the latent structure of the Patient Satisfaction With Cancer Care, and the second dataset was used to validate the said structure. We had a very large sample that facilitated calculation of reliable correlation coefficients for the Patient Satisfaction With Cancer Care. This approach is in accordance with guides on sample sizes for factor analysis/principal components analysis.[27,28] In addition, the principal components analysis (PCA) solutions include many high variables markers and therefore could have facilitated stable and reliable estimates of correlation coefficients with even a smaller sample size.[29] Before conducting the PCA, suitability of the data for dimensionality analysis was assessed using various criteria (eg, examination of the correlation matrix for correlations of.30 and above). The PCA was conducted to reduce the data to a few components that could be more easily described. We performed an initial PCA, using Sample 1 data, without rotation to facilitate extraction and examination of meaningful components, based on eigenvalues and scree plot criteria that more accurately describe the latent structure of the Patient Satisfaction With Cancer Care. The Kaiser-Meyer-Olkin value (KMO), an index of sampling adequacy, was used to determine the suitability of the data for dimensionality analysis.[30,31] In addition, we examined the scree plot of eigenvalues to help determine the number of components to retain. We subsequently rotated the initial factor solution using the VARIMAX technique. Items from Sample 2 were also subjected to a PCA to replicate and test the evidence of the structure of the PCA obtained from Sample 1 through successive unconstrained exploratory procedures. We conducted similar PCA for Sample 2 (N_2) as described above for Sample 1 (N_1).

Measurement reliability analysis

Scale reliability assessment was conducted to determine the degree to which items of the Patient Satisfaction With Cancer Care represent a coherent set that measures the same underlying construct. Cronbach coefficient alpha was used as an index of internal consistency of the Patient Satisfaction With Cancer Care. Measurement reliability analysis was conducted separately for Sample 1 and Sample 2.

RESULTS

The mean age of the analytic sample was 51 years (range, 18-98 years). Most of the sample was female (approximately >80%) and included participants from diverse racial/ethnic backgrounds, including white (43%), black (32%), Hispanic/Latino (23%), Asian (1%), American Indian/Alaska Native (0.5%), and other (0.5%). Half of the sample reported only a high school education or less. Participants presented with abnormal test findings or diagnosis from various types of cancer, including approximately 64% breast, 11% cervix, 12% colorectal, 13% prostate, and 0.5% multiple concurrent cancer sites. Detailed demographic and clinical characteristics of study participants are provided in Table 1. All participants provided informed consent for participation. The institutional review board of all participating institutions approved this study.

Sample 1, N1—Testing of Patient Satisfaction With Cancer Care Latent Structure

Suitability for factor analysis (Sample 1, N1)

Examination of the items correlation matrix revealed the presence of many correlation coefficients of.30 and higher. In addition, the KMO value was 0.95, exceeding the recommended value of 0.60.[30,31] The Bartlett Test of Sphericity also reached statistical significance (chi-square [2378] = 7850.920; $P = .001$), which also supported the appropriateness of dimensionality analyses of the correlation matrix.[32] Values were skewed toward favorable ratings, with a mean coefficient of skewness of 1.45 (range, −2.2 to −0.5).

Table 1. Demographic and Clinical Characteristics of 891 Participants

Characteristic	No.	Mean (SD)
Age, 18-98 y	843	51.43 (13.77)
	No.	**%**
Cancer site		
Breast	572	64.2
Cervix	96	10.77
Colorectal	107	12.01
Prostate	112	12.57
Multiple concurrent cancer sites	4	0.45
Sex		
Female	686	81.28
Male	158	18.72
Race/ethnicity		
White	360	43.22
Black/African American	266	31.93
Asian	9	1.08
American Indian/Alaska Native	4	0.48
Hispanic or Latino	190	22.81
Other	4	0.48
Primary language		
English	740	87.78
Spanish	87	10.32
Other	16	1.9
Birth country		
United States	647	82.32
Other	139	17.68
Marital status		
Single/never married	256	30.51
Married/living as married	339	40.41
Divorced/separated	190	22.65
Widowed	54	6.44
Education		
8th grade or less	69	8.93
Some high school	106	13.71
High school diploma (including equivalency)	196	25.36
Some college/vocational after high school	182	23.54
Associate degree	58	7.5
College graduate	100	12.94
Graduate or professional degree	62	8.02
Household income		
Less than $10,000	219	30.85
$10,000 to $19,999	134	18.87
$20,000 to $29,999	88	12.39
$30,000 to $39,999	69	9.72
$40,000 to $49,999	38	5.35
$50,000 or more	162	22.82
Employment status		
No current employment	443	56.58
Part-time employment	106	13.54
Full-time employment	234	29.89
Health insurance coverage		
Yes	681	83.15
No	138	16.85

SD indicates standard deviation.

Construct validity (Sample 1, N1)

The initial unrotated PCA revealed the presence of 5 components with eigenvalues >1 (λ>1): 12.698, 1.734, 1.383, 1.087, and 1.081, which explained 45.35%, 6.19%, 4.94%, 3.88%, and 3.86% of the total cumulative variance (64.22%), respectively. Inspection of the scree plot revealed a clear break after the second component. Cattell's scree plot test and the eigenvalues criteria suggested that 2 components could be retained for further investigation.[33] The components matrix showed that approximately 82% of the items (the first 23 items) loaded on the first component, with factor or component loadings ranging from 0.51 to 0.86. Of these 23 items, 5 loaded on factors 3 to 5, with component loadings ranging from −0.31 to 0.44. Another set of 5 additional items loaded moderately to strongly on factors 2 to 4, with component loadings ranging from 0.33 to 0.92. This second set of 5 items seems related primarily to time waiting at the hospital, transportation and money concerns, and explication of medical tests and health condition. Subsequently, we removed items with moderate loadings on multiple components because of plausible overlapping contributions. We also decided to not include components defined by just 1 or 2 variables, because such components are unstable, generally account for a very small percentage of the variance, and are difficult to correctly interpret.[34] On the basis of these criteria, we ended up with a 1-dimensional 18-item Patient Satisfaction With Cancer Care measure, as indicated by a single-component structure with items forming a coherent set that explained 62% of the variance in patient satisfaction with cancer-related care (Table 2). The results of our psychometric analyses support the validity of Patient Satisfaction With Cancer Care for this sample.[34,35]

Sample 2, N2—Validation of Patient Satisfaction With Cancer Care Latent Structure

Suitability for Factor Analysis (Sample 2, N2)

We tested the emergent structure of the data in Sample 1 by conducting another PCA on data from Sample 2. This approach is based on the notion that successful replication through successive unconstrained exploratory procedures will substantiate the underlying structure of the Patient Satisfaction With Cancer Care beyond any constrained confirmatory procedure. Similar to Sample 1, examination of the correlation matrix for Sample 2 revealed the presence of many correlation coefficients of .30 and higher. In addition, the KMO value was 0.95, exceeding

Table 2. Component Loadings for Sample 1 ($N_1 = 453$) and Sample 2 ($N_2 = 438$): Correlations Between Individual Items and the Underlying Component

Patient Satisfaction With Cancer Care Scale Items	Component Loadings	
	Eigenvalue (λ) 14.58	Eigenvalue (λ) 15.25
	Sample 1	Sample 2
1. I felt that my health concerns were understood.	0.782	0.756
2. I felt that I was treated with courtesy and respect.	0.762	0.739
3. I felt included in decisions about my health.	0.816	0.751
4. I was told how to take care of myself.	0.741	0.725
5. I felt encouraged to talk about my personal health concerns.	0.758	0.715
6. I felt I had enough time with my doctor.	0.774	0.790
7. My questions were answered to my satisfaction.	0.805	0.815
8. Making an appointment was easy.	0.549	0.577
9. I knew what the next step in my care would be.	0.670	0.745
10. I feel confident in how I deal with the health care system.	0.744	0.791
11. I was able to get the advice I needed about my health issues.	0.817	0.851
12. I knew who to contact when I had a question.	0.695	0.747
13. I received all the services I needed.	0.798	0.780
14. I am satisfied with the care I received.	0.855	0.829
15. The doctors seemed to communicate well about my care.	0.830	0.792
16. I received high-quality care from my regular doctor.	0.723	0.752
17. I received high-quality care from my specialists.	0.811	0.803
18. My regular doctor was informed about the results of the tests I got.	0.541	0.630

Extraction method: principal components analysis.

the recommended value of 0.6.[30,31] The Bartlett Test of Sphericity also reached statistical significance (chi-square [2378] = 7853.56; $P = .001$), supporting the appropriateness of dimensionality analyses of the correlation matrix.[32]

Construct validity (Sample 2, N2)

The initial unrotated PCA revealed the presence of 5 components with eigenvalues >1 (λ>1): 13.12, 1.76, 1.39, 1.20, and 1.03, which explained 46.87%, 6.31%, 4.96%, 4.28%, and 3.66% of the total cumulative variance (66.09%), respectively. Inspection of the scree plot revealed a clear break after the second component. Cattell's 1966 scree plot test and the eigenvalues criteria supported the retention of 2 components for further investigation.[33] Similar to the PCA for Sample 1, the components matrix showed that approximately 82% of the items (the first 23 items) loaded on the first component, with factor or component loadings ranging from 0.48 to 0.86. Of these 23 items, 8 loaded on factors 2, 4, and 5, with component loadings ranging from −0.41 to 0.47. Another set of 5 additional items loaded moderately to strongly on factors 2 to 5, with component loadings ranging from −0.62 to 0.68. Similar to the structure of the Patient Satisfaction With Cancer Care in Sample 1, the second set of 5 items in Sample 2 seemed to involve time waiting at the hospital, transportation and money concerns, and explication of medical tests and health condition. As previously described for Sample 1, we removed items with moderate loadings on multiple components (2 or more) because of issues related to overlapping contribution in Sample 2. Just as in Sample 1, we did not include components defined by just 1 or 2 variables, because such components are unstable, account for a small percentage of the variance, and are difficult to reliably interpret.[34] On the basis of these criteria, we also ended up with an 18-item 1-dimensional measure for Sample 2 as indicated by a 1-component structure (Table 2). Results of our structural analyses supported the use of the Patient Satisfaction With Cancer Care as a valid measure for this sample and more importantly confirmed the underlying structure of the Patient Satisfaction With Cancer Care through successive unconstrained exploratory procedures.[34,35]

Patient Satisfaction With Cancer Care Reliability and Convergent and Divergent Validity

Scale reliability assessment conducted for the 18-item Patient Satisfaction With Cancer Care

Internal consistency—the degree to which items that make up this scale represent a coherent set that measures the same underlying construct—was evaluated using Cronbach coefficient alpha. The results showed Cronbach coefficient alphas of approximately 0.95 and 0.96 based on standardized items for the Patient Satisfaction With Cancer Care for Sample 1 and Sample 2, respectively. The

scale reliability assessment supported the use of the Patient Satisfaction With Cancer Care as a reliable tool of satisfaction with cancer care for this sample.[36]

Convergent and divergent validity

The Patient Satisfaction With Cancer Care total score for Sample 1 ($N_1 = 453$) correlated with subscales of the Communication and Attitudinal Self-Efficacy–Cancer (Understand and Participate in Care [$r = 0.40$, $P = .001$] and Seek and Obtain Information [$r = 0.32$, $P = .004$]). The results, however, did not reveal any statistically significant correlation between the Patient Satisfaction With Cancer Care total score and age, primary language, marital status, and scores on the REALM long form (all P values $>.05$). Likewise, the Patient Satisfaction With Cancer Care total score for Sample 2 ($N_2 = 438$) positively correlated with subscales of the Communication and Attitudinal Self-Efficacy–Cancer: Understand and Participate in Care ($r = 0.51$, $P = .001$), Maintain a Positive Attitude ($r = .30$, $P = .01$), and Seek and Obtain Information ($r = 0.39$, $P = .001$). Again, the analysis revealed no statistically significant correlation between the Patient Satisfaction With Cancer Care total score and age, primary language, or marital status (all P values $>.05$). Convergent and divergent validity analyses examined the degree to which the Patient Satisfaction With Cancer Care correlates with measures that assess related constructs (eg, the Understand and Participate in Care and the Seek and Obtain Information subscales of the Communication and Attitudinal Self-Efficacy–Cancer) and differ from measures or indices of other unrelated constructs (eg, age, primary language, or marital status), hence confirming that the items of the Patient Satisfaction With Cancer Care formed a coherent set that assesses the specific construct of patient satisfaction with the cancer-related care they received.

DISCUSSION

We designed the Patient Satisfaction With Cancer Care to be a simple and easy to administer tool to assess satisfaction with cancer-related care for individuals from diverse cultural and socioeconomic populations. An important goal for developing the Patient Satisfaction With Cancer Care was to ensure that the measure assesses experiences common to all patients regardless of whether they were navigated. This approach is expected to ensure the applicability and relevance of this measure to people from comparable racial, ethnocultural, and socioeconomic backgrounds.

The results of our structural analysis and psychometric validation revealed a parsimonious and reliable 1-component solution for the Patient Satisfaction With Cancer Care. This measure provides a milieu-specific patient-oriented approach for assessing perceived relevance and satisfaction with cancer care for individuals from diverse racial, ethnocultural, and socioeconomic backgrounds. The Patient Satisfaction With Cancer Care demonstrates high construct validity. The degree to which the items of the Patient Satisfaction With Cancer Care constitute a coherent set that assesses the underlying construct of patient satisfaction with cancer care was demonstrated by high indices of internal consistency and reliability.

The Patient Satisfaction With Cancer Care differs from previous generic scales in that it focuses on satisfaction with cancer-related care rather than the broader concept of healthcare in general or the narrower concept of cancer treatment for a particular cancer, disease stage, or location (hospital or ambulatory).[37-40] The Patient Satisfaction With Cancer Care addresses the broad domain of cancer-related care, including diagnostic testing in addition to treatment rather than focusing on particular or specific aspects of cancer care.[41-43]

The limitations of these findings merit comment. First, we adapted and modified items from existing instruments, but we did not conduct cognitive interviewing.[44] However, a pilot study of the questionnaire revealed no problem that would have indicated a need to modify questionnaire items to help improve participants' understanding or interpretation of the items. In addition, the Patient Satisfaction With Cancer Care scale was administered orally to minimize effects of low literacy; therefore, it is not certain that similar results would be obtained from participants who self-administer the scale.

Second, consistent with previous satisfaction measures, we observed significant skewing or a tendency toward the higher end of satisfaction.[45] Whether this represents truly favorable experiences or reflects low expectations is unknown.[3] We did not specifically query patients about expectations. For many patients, their abnormal screening/diagnosis may have been their first experience with cancer-related care. Thus, they may have used a priori general healthcare experiences to form their expectations, which could explain the trend toward the higher end of reported satisfaction. This could also represent a social desirability response bias related to interview format.[46] Further studies are needed to help determined if

this finding will remain if patients respond anonymously and whether this ceiling effect will affect the sensitivity of the scale.

Furthermore, about 80% of the sample were women. Further studies are needed to confirm generalizability of the Patient Satisfaction With Cancer Care to men. Also, the Patient Satisfaction With Cancer Care accounted for 60% of the variance in patient satisfaction. Follow-up studies are needed to identify plausible factors that could account for the unexplained portion of this variance.

Lastly, we did not assess the responsiveness of the measure to change and/or how well it matches clinical impression. That is, we do not know how well the Patient Satisfaction With Cancer Care will capture differences in healthcare processes. Some aspects of care such as interpersonal processes may have a much greater impact on satisfaction than technical aspects.[46-48]

The strengths of the study include psychometric assessment of the Patient Satisfaction With Cancer Care measure with medically underserved and underrepresented individuals from racial/ethnic minorities and lower socioeconomic populations across different types of healthcare systems (eg, community health centers, Veterans Administration, and university- and community-based oncology practices). The development of the Patient Satisfaction With Cancer Care represents an initial attempt to develop and assess the validity and reliability of a context-specific measure of satisfaction with cancer-related care that is applicable to underserved and traditionally underrepresented racial/ethnic minorities and lower income individuals who face a variety of barriers to cancer care.

Validation of this Patient Satisfaction With Cancer Care measure will facilitate examination of the impact of patient navigation on cancer-related care.[12] Further studies should examine the predictive validity of the Patient Satisfaction With Cancer Care for treatment-related outcomes within longitudinal research settings. Our analyses showed divergent and convergent capabilities of the Patient Satisfaction With Cancer Care. Additional studies that examine divergent and convergent characteristics of the Patient Satisfaction With Cancer Care with other relevant psychometrically valid and reliable health measures will provide evidence of the strength of the Patient Satisfaction With Cancer Care and further inform the underlying structure and validity of this measure for cancer patients. This scale, the Patient Satisfaction With Cancer Care, should prove useful for evaluation of patient navigation not only in the participating 9 sites of the NCI funded Patient Navigation Research Program, but in other cancer navigation programs as well.

CONFLICT OF INTEREST DISCLOSURES

Supported by grants from the National Cancer Institute (3U01CA116924-03S1, U01 CA116924-01, 1R25CA 10261801A1, U01CA116892, U01CA 117281, U01CA116903, U01CA116937, U01CA116885, U01CA116875, and U01 CA116925) and the American Cancer Society (SIRSG-05-253-01). The contents are solely the responsibility of the authors and do not necessarily represent the official views of the Center to Reduce Cancer Health Disparities, NCI.

REFERENCES

1. Bredart A, Razavi D, Roberson C, et al. Assessment of quality of care in an oncology institute using information on patients' satisfaction: *Oncology.* 2001;61:120-128.
2. Sanda MG, Dunn RL, Michalski J, et al. Quality of life and satisfaction with outcome among prostate-cancer survivors. *N Engl J Med.* 2008;358:1250–1261.
3. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. *Soc Sci Med.* 1997;45:1829-1843.
4. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
5. Pascoe GC. Patient satisfaction in primary health care: a literature review and analysis. *Eval Program Plann.* 1983;6: 185-210.
6. Cleary PD, McNeil BJ. Patient satisfaction as an indicator of quality of care. *Inquiry.* 1988;25:25-36.
7. Lenderking WR. Brief reflections on treatment satisfaction. *Value Health.* 2005;8:S2-S5.
8. Saila T, Matilla E, Kaila M, et al. Measuring patient assessments of the quality of outpatient care: a systematic review. *J Eval Clin Pract.* 2008;14:148-154.
9. Ware JE Jr, Snyder MK, Wright WR, et al. Defining and measuring patient satisfaction with medical care. *Eval Program Plann.* 1983;6:247-263.
10. Williams B. Patient satisfaction: a valid concept? *Soc Sci Med.* 1994;38:509-516.
11. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. *Eval Program Plann.* 1982;5:233-237.
12. Chow A, Meyer E, Darzi A, Athanassion T. Patient-reported outcome measures: the importance of patient satisfaction in surgery. *Surgery.* 2009;146:435-443.
13. Gold DT, McClung B. Approaches to patient education: emphasizing the long-term value of compliance and persistence. *Am J Med.* 2006;119:S32-S37.
14. Tomlinson JS. Patient satisfaction: an increasingly important measure of quality. *Ann Surg Oncol.* 2006;13:764-765.
15. Borras JM, Sanchez-Hernandez A, Navarro M, et al. Compliance, satisfaction, and quality of life of patients with colorectal cancer receiving home chemotherapy or outpatient treatment: a randomised controlled trial. *BMJ.* 2001;322: 826.

16. Lipscomb J, Reeve BB, Clauser SB, et al. Patient-reported outcomes assessment in cancer trials: taking stock, moving forward. *J Clin Oncol.* 2007;25:5133-5140.
17. Bredart A, Bottomley A. Treatment satisfaction as an outcome measure in cancer clinical trials. *Expert Rev Pharmacoecon Outcomes Res.* 2002;2:597-606.
18. Freund K, Battaglia TA, Calhourn E, et al. National Cancer Institute Patient Navigation Research Program: methods, protocol, and measures. *Cancer.* 2008;113:3391-3399.
19. Liberati A, Confalonierie A, Martino G, et al. Patients' assessment of quality of care: a survey of a group of breast cancer patients in Italy. *Tumori*. 1985;71:491-497.
20. Bredart A, Bottomley A, Blazeby JM, et al. An international prospective study of the EORTC cancer in-patient satisfaction with care measure (EORTC IN-PATSAT32). *Eur J Cancer.* 2005;41:2120-2131.
21. Oberst MT. Methodology in behavioural and psychosocial cancer research: patients' perceptions at care: measurement of quality and satisfaction. *Cancer.* 1984;53(10 suppl):2366-2375.
22. Stewart AL, Ware JE. Measuring Functioning and Well-Being: The Medical Outcomes Study Approach. Durham, NC: Duke University Press; 1992.
23. Wiggers JH, Donovan KO, Redman S, Sanson-Fisher RW. Cancer patient satisfaction with care. *Cancer.* 1990;66:610-616.
24. Lo C, Burman LD, Rodin G, Zimmermann C. Measuring patient satisfaction in oncology palliative care: psychometric properties of the FAMCARE-patient scale. *Qual Life Res.* 2009;18:747-752.
25. Teno JM, Lima JC, Lyons KD. Cancer patient assessment and reports of excellence. Reliability and validity of advanced cancer patient perceptions of the quality of care. *J Clin Oncol.* 2009;27:1621-1626.
26. Wolf MS, Chang CH, Davis TC, Makoul GT. Development and validation of the Communication and Attitudinal Self-Efficacy (CASE) scale: Cancer version. *Patient Educ Couns.* 2005;57:333-341.
27. Comrey AL, Lee HB. A First Course in Factor Analysis. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1992.
28. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 5th ed. Boston, MA: Allyn and Bacon; 2007.
29. Guadagnoli, Edward, Wayne F. Velicer. Relation of sample size to the stability of component patterns. *Psychol Bull.* 1988;103:265-275.
30. Kaiser H. A second generation Little Jiffy. *Psychometrika.* 1970;35:401-415.
31. Kaiser H. An index of factorial simplicity. *Psychometrika.* 1974;39:31-36.
32. Bartlett MS. A note on the multiplying factors for various chi square approximations. *J R Stat Soc Series B Stat Methodol.* 1954;16:296-298.
33. Cattell RB. The scree test for the number of factors. *Multivariate Behav Res.* 1966;1:245-276.
34. Tabachnick BG, Fidell LS. Using multivariate statistics. 5th ed. Boston, MA: Allyn and Bacon; 2007.
35. Thurstone LL. Multiple factor analysis. Chicago, IL: University of Chicago Press; 1947.
36. Nunnally JO. Psychometric Theory. New York, NY: McGraw-Hill; 1978.
37. Atkinson MJ, Sinha A, Hass SL, et al. Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease. *Health Qual Life Outcomes.* 2004;2:12.
38. Bredart A, Razuvi D, Robertson C, et al. Comprehensive assessment of satisfaction with care: preliminary psychometric analysis in French, Polish, Swedish and Italian oncology patients. *Patient Educ Couns.* 2001;43:243-253.
39. Greenfield TK, Attkisson CC. Steps toward a multifactorial satisfaction scale for primary care and mental health services. *Eval Program Plann.* 1989;12:271-278.
40. Sulmasy DP, McIlvane JM, Pasley PM, Rahn M. A scale for measuring patient perceptions of the quality of end-of-life care and satisfaction with treatment: the reliability and validity of QUEST. *J Pain Symptom Manage.* 2002;23:458-470.
41. Defossez G, Mathouwin-Pelissier S, Ingrand I, et al. Satisfaction with care among patients with nonmetastatic breast cancer: development and first steps of validation of the Defossez G. REPERES-60 questionnaire. *BMC Cancer.* 2007;7:129.
42. Llewellyn CD, Robert H, McGurk M, Weinman J. Development and preliminary validation of a new measure to assess satisfaction with information among head and neck cancer patients: the satisfaction with cancer information profile (SCIP). *Head Neck.* 2006;28:540-548.
43. Lipscomb J, Gotay CC, Snyder CF. Patient-reported outcomes in cancer: a review of recent research and policy initiatives, *CA Cancer J Clin*. 2007;57:278-300.
44. Drennan J. Cognitive interviewing: verbal data in the design and pretesting of questionnaires. *J Adv Nurs.* 2003;42:57-63.
45. Crow R, Gage H, Hampson S, et al. The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. *Health Technol Assess.* 2002;6:1-244.
46. Chang JT, Hays RD, Shekelle PG, et al. Patients' global ratings of their health care are not associated with the technical quality of their care. *Ann Intern Med.* 2006;144:665-672.
47. Napoles AM, Gregorich SE, Santoyo-Olsson J, O'Brien H, Stewart AL. Interpersonal processes of care and patient satisfaction: do associations differ by race, ethnicity, and language? *Health Serv Res.* 2009;44:1326-1344.
48. Rao M, Clarke A, Sanderson C, Hammersley R. Patients' own assessments of quality of primary care compared with objective records based measures of technical quality of care: cross sectional study. *BMJ.* 2006;333:19.

Validity of Selected Patient Safety Indicators: Opportunities and Concerns

Haytham MA Kaafarani, MD, MPH, Ann M Borzecki, MD, MPH, Kamal MF Itani, MD, FACS, Susan Loveland, MAT, Hillary J Mull, MPP, Kathleen Hickson, RN, MN, Sally MacDonald, RNC, Marlena Shin, JD, MPH, Amy K Rosen, PhD

BACKGROUND: The Agency for Healthcare Research and Quality (AHRQ) recently designed the Patient Safety Indicators (PSIs) to detect potential safety-related adverse events. The National Quality Forum has endorsed several of these ICD-9-CM-based indicators as quality-of-care measures. We examined the positive predictive value (PPV) of 3 surgical PSIs: postoperative pulmonary embolus and deep vein thrombosis (pPE/DVT), iatrogenic pneumothorax (iPTX), and accidental puncture and laceration (APL).

STUDY DESIGN: We applied the AHRQ PSI software (v.3.1a) to fiscal year 2003 to 2007 Veterans Health Administration (VA) administrative data to identify (flag) patients suspected of having a pPE/DVT, iPTX, or APL. Two trained nurse abstractors reviewed a sample of 336 flagged medical records (112 records per PSI) using a standardized instrument. Inter-rater reliability was assessed.

RESULTS: Of 2,343,088 admissions, 6,080 were flagged for pPE/DVT (0.26%), 1,402 for iPTX (0.06%), and 7,203 for APL (0.31%). For pPE/DVT, the PPV was 43% (95% CI, 34% to 53%); 21% of cases had inaccurate coding (eg, arterial not venous thrombosis); and 36% featured thromboembolism present on admission or preoperatively. For iPTX, the PPV was 73% (95% CI, 64% to 81%); 18% had inaccurate coding (eg, spontaneous pneumothorax), and 9% were pneumothoraces present on admission. For APL, the PPV was 85% (95% CI, 77% to 91%); 10% of cases had coding inaccuracies and 5% indicated injuries present on admission. However, 27% of true APLs were minor injuries requiring no surgical repair (eg, small serosal bowel tear). Inter-rater reliability was >90% for all 3 PSIs.

CONCLUSIONS: Until coding revisions are implemented, these PSIs, especially pPE/DVT, should be used primarily for screening and case-finding. Their utility for public reporting and pay-for-performance needs to be reassessed. (J Am Coll Surg 2011;212:924–934. © 2011 by the American College of Surgeons)

Since publication of the Institute of Medicine's 2 landmark reports, "To Err is Human" and "Crossing the Quality Chasm" in 2000, patient safety has become a national health care priority.[1,2] The number of studies aimed at measuring adverse events, safety-related events, errors, or patient harm has increased exponentially in the last few years. Medical record review promises richer clinical detail in identifying adverse events, but is resource-intensive and expensive.[3] Traditional mortality and morbidity conferences show less than 25% sensitivity in detecting adverse events when compared with the large chart-based quality measurement initiatives such as the National Surgical Quality Improvement Program (NSQIP).[4,5]

Recently, in an effort to improve and increase detection of potentially preventable safety events in acute-care hospitals, the Agency for Healthcare Research and Quality (AHRQ) designed a set of evidence-based ICD-9-CM-based algorithms called Patient Safety Indicators (PSIs). The PSIs represent a significant advance in the field of patient safety because they use readily available hospital discharge data that can be risk adjusted to screen for safety-related events in the inpatient setting.[6] The PSIs were ini-

Disclosure Information: Nothing to disclose.

Abstract presented American College of Surgeons 95th Annual Clinical Congress, Surgical Forum, Chicago, IL, October 2009.

Received April 19, 2010; Revised July 8, 2010; Accepted July 8, 2010.

From the Department of Surgery, Tufts Medical Center, Tufts University School of Medicine (Kaafarani), VA Boston Healthcare System and the Center for Health Quality, Outcomes and Economic Research (Kaafarani), Boston University School of Public Health (Borzecki, Mull, Rosen), Boston University School of Medicine (Borzecki, Itani), Harvard Medical School (Itani), and the Center for Organization, Leadership and Management Research, VA Boston Healthcare System (Loveland, Mull, Shin, Rosen), Boston, MA; the Center for Health Quality, Outcomes and Economic Research, Bedford VAMC, Bedford, MA (Borzecki, Hickson, MacDonald); and the Department of Surgery, VA Boston Healthcare System, West Roxbury, MA (Itani, Hickson).

Correspondence address: Amy K Rosen, PhD, Center for Organization, Leadership and Management Research (COLMR), VA Boston Healthcare System, 150 S Huntington Ave, Boston MA 02130. email: akrosen@bu.edu or amy.rosen2@va.gov

ISSN 1072-7515/11/$36.00
doi:10.1016/j.jamcollsurg.2010.07.007

Abbreviations and Acronyms

AHRQ = Agency for Healthcare Research and Quality
APL = accidental puncture and laceration
CMS = Center for Medicare and Medicaid Services
iPTX = iatrogenic pneumothorax
IRR = inter-rater reliability
pPE/DVT = postoperative pulmonary embolus and deep vein thrombosis
PPV = positive predictive value
PSI = Patient Safety Indicator
VA = Veterans Health Administration
VTE = venous thromboembolism

tially intended as screening tools to identify safety-related events or as case-finding tools for internal quality improvement purposes. However, in the last several years, they have been used by multiple organizations for hospital profiling and pay-for-performance purposes.[7,8] Eight of the PSIs have been endorsed by the National Quality Forum (NQF) as hospital performance measures, and 4 (plus 1 composite measure) have been adopted by the Centers for Medicare and Medicaid Services (CMS) for hospital comparisons of quality and safety and financial reimbursement.[9,10]

With the increased use of PSIs for quality assessment, public reporting, hospital profiling, and reimbursement,[8,11] many clinicians, policymakers, and researchers have raised concerns over the validity of PSIs, given that the PSI algorithms are based on administrative data that are well known for their variability and inconsistency in the coding of diagnoses and procedures. Some of these coding issues include ambiguity in ICD-9-CM coding guidelines, variation in coding practices across different hospitals, and the codes' inability to differentiate between events that happened de novo versus those that were present on admission.[12,13] Although the current literature suggests that PSIs examined to date have moderate to high sensitivities and specificities, their measured performance has been found to depend on the nature of the specific PSI, the nature of the adverse events it targets, and the method of validation used to test its performance. In these multiple validation attempts, the calculated positive predictive values (PPVs) of PSIs against chart abstraction as the "gold standard," ranged between 44% and 91%.[14-20]

Although the criterion validity of several PSIs against the gold standard has been examined in the private sector, little is known about whether these results are reliable in other health care settings, such as the Veterans Health Administration (VA). In this study, we examined the positive PPV of 3 surgical PSIs against medical record review: postoperative pulmonary embolus and deep vein thrombosis (pPE/DVT), iatrogenic pneumothorax (iPTX), and accidental puncture and laceration (APL). These 3 PSIs have all been endorsed by the National Quality Forum as quality measures; APL and iPTX have been adopted by CMS.

METHODS

We applied the AHRQ PSI software (v.3.1a) to VA fiscal year 2003 to 2007 (October 1, 2002 to September 30, 2007) administrative data from a sample of 28 VA hospitals to identify (flag) patients suspected of having a pPE/DVT, iPTX, or APL. To determine rates of true and false positives for each PSI, trained nurses conducted a retrospective chart review of 336 flagged charts (112 charts per PSI) using standardized chart abstraction tools and guidelines developed by the AHRQ and modified for the VA's electronic medical record. Inter-rater reliability testing of the 2 nurse abstractors was also performed. The required Institutional Review Board (IRB) approvals from the Bedford VA Medical Center and the VA Boston Healthcare System were obtained to conduct this study.

PSI overview and definitions

PSIs are calculated as rates: the numerator consists of all discharges with diagnoses, procedures, and/or other attributes indicating the occurrence of the particular adverse event; the denominator includes all discharges at risk for the adverse event, with certain exclusions. All denominators include patients aged 18 and older and exclude patients with Major Diagnostic Category (MDC) 14 (pregnancy, childbirth, and puerperium).[21]

Postoperative PE/DVT detects the occurrence of postoperative venous thromboembolism (VTE) by identifying discharges that have any of the ICD-9-CM codes for PE or DVT in a secondary diagnosis field. The denominator includes all surgical discharges as defined by specific Diagnosis-Related Groups (DRGs) and an ICD-9-CM code for an operating room procedure, narrowing the denominator to those cases most likely to be preventable. It excludes all patients with pre-existing (principal diagnosis or secondary diagnosis present on admission, if known) PE or DVT and patients who underwent a procedure to interrupt the vena cava either as the only operating room procedure during admission or before or on the same day of the index operating room procedure.

The occurrence of iPTX is targeted by identifying discharges with the ICD-9-CM code 512.1 in any secondary diagnosis field. The denominator includes all surgical and medical discharges of patients with specific DRGs. It excludes patients with the code 512.1 in a primary diagnosis field; a code for diaphragmatic surgery repair; a cardiac surgery DRG; a thoracic surgery or lung or pleural biopsy code; and a diagnosis code for chest trauma or pleural ef-

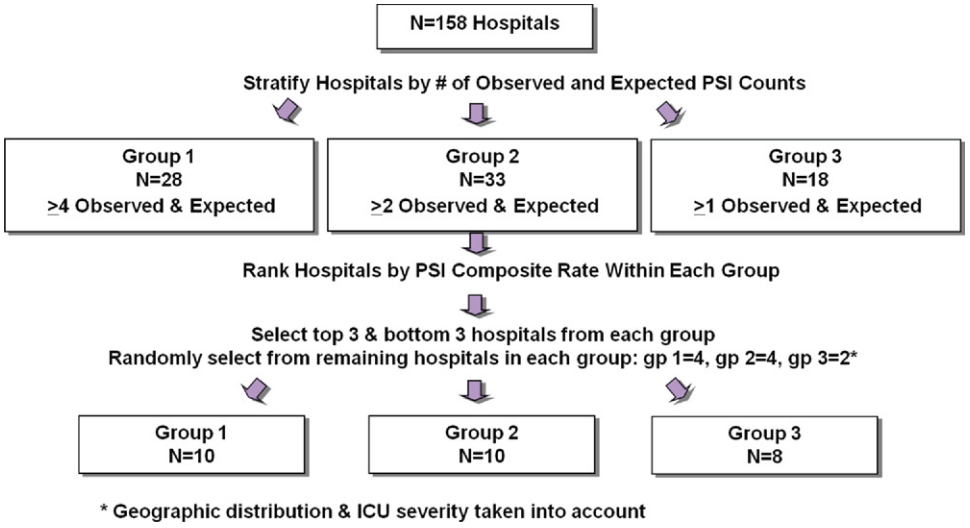


Figure 1. Hospital sampling strategy. PSI, Patient Safety Indicator.

fusion. Iatrogenic pneumothoraces caused by procedures performed within 72 hours before admission are included. Similar to pPE/DVT, the denominator was purposely restrictive in order to capture the most "preventable" cases.

APL captures the occurrence of accidental cut, puncture, perforation, or hemorrhage during medical care (ICD-9-CM codes E870.0 through E870.9) or accidental puncture or laceration during a procedure (ICD-9-CM code 998.2) in a secondary diagnosis field. The denominator includes all surgical and medical discharges excluding patients whose APL was present on admission or in a primary diagnosis field.

Hospital selection

Our initial hospital sample included 158 acute-care VA hospitals. To obtain a manageable number of hospitals for chart review, and to minimize variation in coding across hospitals, we selected a sample of hospitals from the 158 that represented a broad spectrum of PSI rates. We grouped the 158 hospitals into 3 tiers based on their observed and expected rates of PSIs, exclusive of PSI 5 (foreign body left during procedure) and PSI 8 (postoperative hip fracture), both of which had low incidence rates across most hospitals. The expected number of PSI events of a specific facility was calculated as the national VA PSI rate multiplied by the PSI denominator of that specific facility. The first group of hospitals included facilities that had a numerator of at least 4 safety-related events in both the expected and observed numerators of each PSI. The second group had at least 2 safety-related events in the expected and observed numerators of each PSI. The third group had at least 1 safety-related event in the expected and observed numerators of each PSI. Hospitals with less than 1 safety-related event in the expected and observed numerators in any of the PSIs were excluded, yielding a final sample of 79 hospitals. Within each tier of hospitals, facilities were ranked using the AHRQ PSI composite rate.[21] The top 3 and bottom 3 hospitals were included in the hospital sample. We then randomly selected from the remaining hospitals within each stratum to obtain a sample of 28 hospitals. To assure balanced geographic representation, 3 hospitals were replaced by the next hospital in rank, for a final sample of 28 hospitals representing diverse geographic regions of the US. The hospital selection process is shown in Figure 1.

Case selection

Four flagged medical records per PSI (pPE/DVT, iPTX, APL) were randomly selected from each of the 28 hospitals for a total of 336 medical records. Based on previously reported PPV estimates, 112 cases per PSI were selected to ensure reasonably narrow PPV confidence intervals (range 10% to 20%). When a certain facility had less than 4 flagged records for a specific PSI, flagged records were reviewed from the next facility within the stratum, for a maximum of 8 records total per hospital.

Medical record abstraction

Two trained nurses used standardized abstraction instruments to review medical records for occurrence of a safety-related event (pPE/DVT, iPTX, or APL); demographics, comorbidities, and risk factors of the patient population; clinical circumstances surrounding the safety-related event; and patient outcomes after the event. Nurse training included several sessions discussing the rationale behind each PSI, the likely sources of information needed from the electronic medical record, and a systematic chronology for

chart abstraction. When the nurse abstractors were uncertain about any item, the medical record was referred to the research team physicians (HK, AB, KI) for resolution.

Inter-rater reliability

After completion of nurse training, we initiated inter-rater reliability (IRR) testing, as recommended in the literature, to obtain a standardized and reliable method of abstraction.[22] At least 10% of the medical records were reviewed by both nurses, and IRR was measured as the percentage of agreement on select key questions of each abstraction tool, such as ascertainment of the safety-related event. Records were abstracted in groups of 5 until >90% agreement was obtained; this typically occurred after the first or second round of IRR, and then nurses were instructed to proceed by abstracting separate charts. After each round of IRR (including the round achieving >90%), a discussion of disagreements took place in the presence of the clinical physicians (HK, AB, KI) with resulting instrument revisions and/or guideline clarifications as appropriate. An additional round of IRR was performed on 5 records toward the end of the abstraction process to check for potential abstractor drift.

Positive predictive value

For each of the 3 PSIs, we calculated PPV as the rate of true positives divided by the number of medical records reviewed and derived 95% confidence intervals for that estimate.

True positive analysis

For patients with a confirmed pPE/DVT, iPTX, or APL, we performed descriptive analyses of multiple continuous and categorical variables including demographics (age, gender, and race or ethnicity), comorbidities, relevant risk factors, and the nature of the surgical procedure and outcomes. For example, for pPE/DVT, the location of the DVT and the index procedure preceding the PE or DVT were identified. For iPTX, we examined the cause of the pneumothorax (eg, subclavian or internal jugular central line insertion, ventilator-associated barotrauma) and the training level of the staff involved in the procedure (attending vs trainee). For APL, we examined the nature of the surgical procedure, the type of puncture or laceration (eg, bowel injury, durotomy, and splenic injury), and the training level of the staff primarily involved in the procedure. All statistical analyses were performed using SAS version 9.1.

False positive analysis

All false positive cases underwent further detailed review to better understand why they were incorrectly flagged by the PSI algorithms.

RESULTS

Of 2,343,088 admissions, 6,080 were flagged for pPE/DVT (0.26%), 1,402 for iPTX (0.06%), and 7,203 for APL (0.31%).

Postoperative pulmonary embolus or deep vein thrombosis

Positive predictive value

Of 112 cases, 48 were true events of postoperative PE or DVT, yielding a PPV of 43% (95% CI 34% to 53%). IRR between chart abstractors was measured at 94%.

True positive analysis

As shown in Table 1, the patient population was entirely male, with a mean age of 70 years. Seventy-one percent were white, non-Hispanic patients. The mean number of comorbidities per patient was 1.6: 17% of the patients were diabetic, 17% had chronic pulmonary disease, 13% had a malignancy, and 8% had congestive heart failure. Twenty-two of the 48 true pPE/DVT patients had only a DVT diagnosis (46%), 19 patients had only a PE diagnosis (40%), and 7 patients had both a DVT and a PE diagnosed (15%). Of the total of 29 DVTs, 23 involved the lower extremity (79%), and only 4 involved the upper extremities (14%). Orthopaedic and abdominal procedures accounted for more than half of the index procedures preceding the VTE events (38% and 21%, respectively); 67% of the index procedures were elective in nature. The median length of stay for these admissions was 22 days. All-cause in-hospital mortality was 19%.

False positive analysis

Of a total of 64 false positive cases, 16 (25%) patients had a PE and/or DVT diagnosis present on admission (diagnosed 6 months or less before admission), 10 patients (16%) had a remote history of PE/DVT (diagnosed more than 6 months before admission), and 14 patients (22%) were diagnosed after admission but before the index procedure. Coding-related inaccuracies accounted for the remaining 24 false positives (38%) (Fig. 2). These included cases of arterial (not venous) thrombosis, superficial (not deep) vein thrombosis or thrombophlebitis, and cases in which a postoperative PE and DVT workup was negative or the etiology for patient postoperative mortality was uncertain (ie, "rule out" PE). Ten of the cases (9%) were considered false positives for diverse reasons (the miscellaneous category). For example, one of the false positive cases had a discharge summary stating: "patient with history and PE compatible with acute appendicitis." It seems likely that, in this instance, the coder mistook the abbreviation "PE" to stand for pulmonary embolus rather than physical examination.

Table 1. Analysis of True Positives for Postoperative Pulmonary Embolus or Deep Vein Thrombosis

Variable	Data
n	48
Demographics	
Age, y, mean (SD)	70 (11)
Gender, male, n (%)	48 (100)
Race/ethnicity, n (%)	
White, non-Hispanic	34 (71)
African American, non-Hispanic	5 (10)
Hispanic	3 (6)
Other/missing	6 (13)
Body mass index, kg/m^2, mean (SD)	27.1 (4.7)
Comorbidities	
No. of comorbidities, mean (SD)	1.6 (1.3)
No. of comorbidities, median (IQR)	1 (1.5 [1–2.5])
Specific comorbidities, n (%)	
Diabetes mellitus	8 (17)
Chronic pulmonary disease	8 (17)
Congestive heart failure	4 (8)
Renal failure	3 (6)
Solid malignancy without metastasis	3 (6)
Metastatic malignancy	2 (4)
Lymphoma	1 (2)
Obesity	1 (2)
Paralysis	1 (2)
Other neurologic disorders	2 (4)
Additional risk factors, n (%)	
Hypercoagulable state	2 (4)
Baseline inability to ambulate	5 (10)
History of spinal cord injury	1 (2)
History of recent trauma	0 (0)
Central venous catheter insertion	
Femoral	1 (2)
Subclavian	5 (10)
Internal jugular	7 (15)
Peripherally inserted central line	13 (27)
Ventilator dependence	8 (17)
Chemotherapy	2 (4)
Use of potentially procoagulant medication	0 (0)
Transfusion of blood or blood products	0 (0)
Index procedures, n (%)	
Elective procedure	32 (67)
Procedure type	
Orthopaedic	18 (38)
Abdominal	10 (21)
Urologic (including nephrectomy)	6 (13)
Thoracic/pulmonary	6 (13)
Cardiac	3 (6)
Neurosurgical	3 (6)
Vascular	2 (4)
Type of venous thromboembolism, n (%)	
Pulmonary embolus only	19 (40)
Deep vein thrombosis only	22 (46)
Both pulmonary embolus and deep vein thrombosis	7 (15)
Location of deep vein thrombosis, when applicable, n= 29, n (%)	
Lower extremity	23 (79)
Upper extremity	4 (14)
Internal jugular	1 (3)
Undocumented location	1 (3)
Outcomes	
Length of stay, d	
Mean (SD)	35 (32)
Median (range)	22 (1–127)
All-cause mortality during admission, n (%)	9 (19)

Numbers might not add to totals due to rounding.
IQR, interquartile range.

Iatrogenic pneumothorax

Positive predictive value

Out of 112 cases reviewed, 82 were true iatrogenic pneumothoraces, yielding a PPV of 73% (95% CI, 64% to 81%]. IRR between the chart abstractors was measured at 94%.

True positive analysis

As shown in Table 2, the population was 94% male and 62% white non-Hispanic; mean age was 68 years, and the mean body mass index was 24.5 kg/m^2. Hypertension, airway lung disease (eg, asthma, COPD), and diabetes mel-

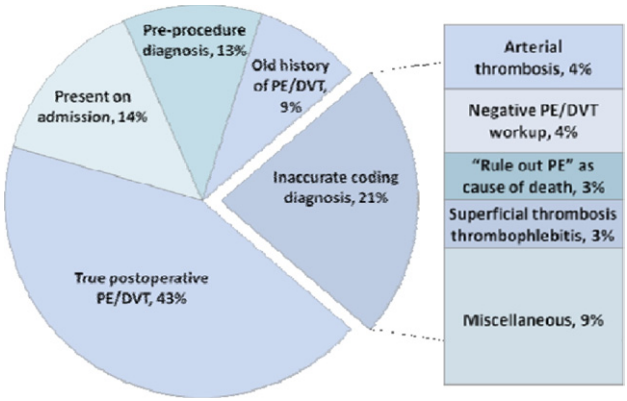


Figure 2. Positive predictive value and analysis of false positives of postoperative pulmonary embolus or deep vein thrombosis (pPE/DVT). Numbers might not add to totals due to rounding. The percentages reported in the figure refer to the percentage of the total number of cases; those reported in the text of the manuscript refer to the percentage of the false positive cases only.

Table 2. Analysis of True Positives for Iatrogenic Pneumothorax

Variable	Data
n	82
Demographics	
Age, y, mean (SD)	68 (14)
Gender, male, n (%)	77 (94)
Race/ethnicity, n (%)	
White, non-Hispanic	51 (62)
African American, non-Hispanic	12 (15)
Hispanic	4 (5)
Other/missing	15 (18)
Body mass index, kg/m^2, mean (SD)	24.5 (6.5)
Comorbidities	
No. of comorbidities, mean (SD)	1.7 (1.3)
No. of comorbidities, median (IQR)	2 (1 [1–2])
Specific comorbidities, n (%)	
Hypertension	31 (38)
Diabetes mellitus	19 (23)
Chronic pulmonary disease	19 (23)
Congestive heart failure	12 (15)
Weight loss	6 (7)
Alcohol abuse	4 (5)
Liver disease	2 (2)
Metastatic malignancy	2 (2)
Peripheral vascular disease	2 (2)
Pulmonary-related risk factors, n (%)	
Airway disease (eg, COPD, asthma)	24 (29)
Respiratory tract malignancy	9 (11)
Infectious lung disease (eg, pneumonia)	8 (10)
Pleural effusion	7 (8)
Interstitial lung disease (eg, lung fibrosis)	3 (4)
Other respiratory disease	5 (6)
Description of the iatrogenic pneumothorax	
Causes of the pneumothorax, n (%)*	
Central venous catheter insertion	31 (38)
Transthoracic needle aspiration (or biopsy)	15 (18)
Cardiac pacemaker placement	11 (13)
Implantable defibrillator insertion	4 (5)
Mechanical ventilation (barotrauma)	3 (4)
Liver biopsy/liver lesion radiofrequency ablation	3 (4)
Cardiopulmonary resuscitation	1 (1)
Miscellaneous procedures near chest/neck	14 (17)
Level of training of person performing procedure, n (%), n = 85	
Attending	28 (33)
Physician-in-training	38 (45)
Unknown	19 (22)
Outcomes	
Length of stay, d	
Mean (SD)	18 (21)
Median (range)	12 (1–140)
All-cause mortality during admission, n (%)	15 (18)

Numbers might not add to totals due to rounding.
*Three patients had 2 potential causative procedures each.
IQR, interquartile range.

litus were the most prevalent risk factors in the patient population (38%, 29%, and 23%, respectively). Of 82 cases of true iatrogenic pneumothoraces, 31 were caused by central venous catheter insertions (38%), 15 by transthoracic needle aspirations or biopsies (18%), 15 by cardiac pacemaker or defibrillator placement (18%), 3 by liver biopsies or liver lesion radiofrequency ablation (4%), 3 by mechanical ventilation (4%), and 1 by cardiopulmonary resuscitation (1%). Several miscellaneous procedures (eg, brachial, axillary or intercostal nerve blocks, pericardiocentesis) accounted for 14 pneumothoraces (17%). Forty-five percent of the pneumothoraces occurred during procedures performed by physicians in training; 33% occurred during procedures performed by attending-level physicians alone. The median length of hospital stay for these patients was 12 days, and the all-cause in-hospital mortality was 18%.

False positive analysis

There were 30 false positive cases, including 10 patients with an old history of pneumothorax or pneumothorax present on admission (more than 72 hours before admission), and 20 cases of inaccurate coding (Fig 3). The latter included cases in which the pneumothorax was spontaneous but not iatrogenic; cases in which lung consolidation or collapse due to pneumonia or empyema were mistaken for iatrogenic pneumothoraces; and cases in which the procedure was known a priori to breach the pleural cavity, such as a thoracic procedure or a lung biopsy.

Accidental puncture or laceration

Positive predictive value

Ninety-five of 112 cases represented "true" cases of accidental punctures or lacerations, yielding a PPV of 85% (95% CI, 77% to 91%). IRR between the chart abstractors was measured at 97%.

True positive analysis

As illustrated in Table 3, the population was 96% male and 63% white non-Hispanic; mean age was 67 years, and mean body mass index was 27.8 kg/m^2. The mean number of comorbidities was 1.5, with the most common comorbidities being hypertension (57%), diabetes mellitus (22%), and chronic pulmonary disease (12%). Seventy-nine of the 95 true iatrogenic events involved the chest or the abdomen (83%); 75 of 95 occurred in the operating room (79%). Injury to vascular structures (inadvertent injury to large arteries or veins), bowel (eg, iatrogenic enterotomy), abdominal organs (eg, splenic injury in a colectomy), genitourinary organs (eg, ureteral injury), and spinal dura (eg, durotomy) accounted for 81% of the punctures or lacerations (23%, 17%, 15%, 13%, and 11%,

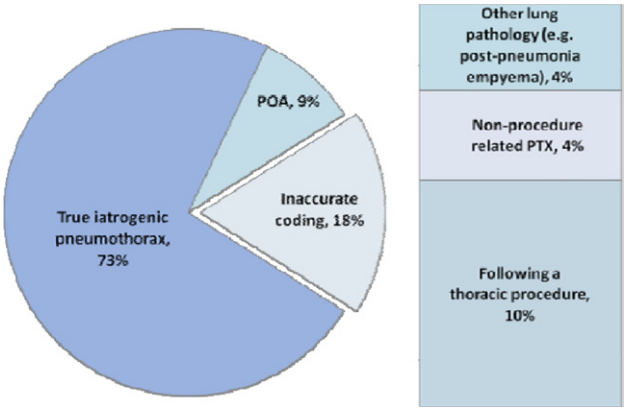


Figure 3. Positive predictive value and analysis of false positives of iatrogenic pneumothorax (iPTX). Numbers might not add to totals due to rounding. The percentages reported in the figure refer to the percentage of the total number of cases; those reported in the text of the manuscript refer to the percentage of the false positive cases only. POA, present on admission.

respectively). Simple serosal bowel tears accounted for 6% of the injuries. When management of these injuries was examined, 26% of them required no surgical management at all (eg, a small serosal bowel tear that was not repaired or a small spinal durotomy that was managed using gelfoam only). Forty-four percent of these injuries occurred during procedures performed by physicians in training; 37% occurred during procedures performed by attending-level physicians alone. In 25% of cases, surgeons reported presence of adhesions or scar tissue; an additional 5% were associated with abnormal anatomy. The median length of stay for patients who sustained these injuries was 8 days. The all-cause in-hospital mortality of these patients was 5%.

False positive analysis

Out of a total of 17 false positive cases, 6 were patients with punctures or lacerations that were present on admission (35%) and 11 were cases of inaccurate coding diagnosis (65%) (Fig. 4). Half of the cases in the latter category were due to the puncture or laceration being nonaccidental (eg, ampulla of Vater "slit" during an endoscopic retrograde cholangiopancreatography to release impacted stones; spontaneous nonprocedure sigmoid colon perforation). The other half involved miscellaneous conditions (eg, suspicion of an intubation-related pharyngeal injury that was later ruled out), including cases in which the reason for coding as an accidental puncture or laceration remained unclear to the chart abstractors.

DISCUSSION

This is the first study to examine the validity of pPE/DVT, iPTX, and APL in the VA. We found that the PPVs of these 3 PSIs varied considerably. Postoperative PE/DVT had a low PPV (43%), iPTX had a moderate PPV (74%), and APL had a moderate-to-high PPV (85%). However, 26% of the accidental punctures or lacerations detected by the APL algorithm were minor injuries that did not require any surgical repair or intervention (eg, serosal bowel tears or dural tears that heal with simple observation or with the use of sealants). Our PPV findings are consistent with those from private sector studies.[14-20,23] Two recent studies estimated the PPV of pPE/DVT to be between 44% and 55%.[14] A separate study estimated the PPV of APL to be around 91%, but similarly noted that 24% of the injuries were "inconsequential," "expected to heal without repair," and for which "the risk may have been acceptable relative to the goals of the procedure."[16] Another recent study estimated the PPV of iPTX to be around 78%, similar to our findings, with 44% of the pneumothoraces resulting from central venous line insertions.[23]

Policy implications

The PSIs we studied show promise as screening tools that can be used to detect patient safety events related to postoperative VTE, iatrogenic pneumothoraces, and accidental punctures or lacerations. Their accuracy can be improved by strategies focused on adjustment of coding guidelines and efforts aimed at educating coders and ameliorating their clinical knowledge. In addition, introduction of "present on admission" flags in administrative data, an intervention already adopted in some states and in many hospitals as of 2007, will clearly improve the PPV of these PSIs by differentiating between new onset diagnoses and those that patients have before admission to the hospital. The ICD-9-CM codes used in the PSI algorithms were initially designed for billing purposes; therefore, using them for clinical and quality of care purposes may require modifications directly to the coding schemes. For example, coding all the "rule out" diagnoses referred to in a physician's note when a postoperative patient experiences shortness of breath (eg, myocardial infarction, pulmonary embolus, pulmonary edema, COPD exacerbation, etc) might make sense from a financial perspective; however, it serves to decrease the specificities and alter the predictive values of the PSIs, which rely heavily on the clinical accuracy of the ICD-9-CM codes. Despite the relative lack of financial incentives in coding in the VA compared with other settings, the fact that our PPVs were similar to or lower than those in the private sector suggests that the shortcomings of the PSIs are inherent to the coding algorithms and practices, and not specific to any one setting. Shifting coding practices to rely on standard accepted clinical criteria rather than physician notes and reports alone will definitely improve the predictive value of PSIs. More importantly, tar-

Table 3. Analysis of True Cases of Accidental Punctures or Lacerations

Variable	Data
n	95
Demographics	
Age, y, mean (SD)	67 (10)
Gender, male, n (%)	91 (96)
Race/ethnicity, n (%)	
White, non-Hispanic	60 (63)
African American, non-Hispanic	11 (12)
Hispanic	6 (6)
Other/missing	18 (19)
Body mass index, kg/m^2, mean (SD)	27.8 (5.5)
Comorbidities	
No. of comorbidities, mean (SD)	1.5 (1.2)
No. of comorbidities, median (IQR)	1 (2 [0–2])
Specific comorbidities, n (%)	
Hypertension	54 (57)
Diabetes mellitus	21 (22)
Chronic pulmonary disease	11 (12)
Congestive heart failure	7 (7)
Liver disease	2 (2)
Alcohol abuse	2 (2)
Metastatic malignancy	3 (3)
Weight loss	5 (5)
Peripheral vascular disease	7 (7)
Description of puncture or laceration	
Location of puncture or laceration	
Head, n (%)	2 (2)
Neck, n (%)	9 (9)
Chest, n (%)	19 (20)
Abdomen, n (%)	60 (63)
Upper extremity, n (%)	1 (1)
Lower extremity, n (%)	4 (4)
Type of puncture or laceration	
Vascular	24 (23)
Gastrointestinal tract	18 (17)
Abdominal organs	16 (15)
Genitourinary tract	14 (13)
Spine dura	11 (11)
Serosal bowel tear	6 (6)
Pleural injury	5 (5)
Heart/lungs	4 (4)
Miscellaneous	6 (6)
Setting where puncture or laceration occurred	
Operating room	75 (79)
Cardiac catheterization suite	4 (4)
Emergency room	1 (1)
Radiology suite	2 (2)
Patient bedside	1 (1)
Other/undetermined	12 (13)
Level of training of person performing procedure, n (%)	
Attending physician	35 (37)
Physician-in-training	42 (44)
Physician assistant	1 (1)
Unknown	17 (18)
Risk factors present at the time of procedure	
Lysis of adhesions	19 (20)
Presence of "scar" tissue	5 (5)
Abnormal anatomy	5 (5)
No. of accidental punctures or lacerations, n (%)	
1	74 (78)
2	14 (17)
≥3	6 (6)
Unknown	1 (1)
Outcomes	
Length of stay, d	
Mean (SD)	15 (22)
Median (range)	8 (1-168)
All-cause mortality during admission, n (%)	5 (5)

Numbers might not add to totals due to rounding.
IQR, interquartile range.

geted coder education is essential to avoid many of the shortcomings of administrative data in general and the PSIs in specific. Misunderstanding of the nuances of clinical and surgical care was evident throughout this project, where physicians' phrases were occasionally taken out of context, such as with the interpretation of PE as "pulmonary embolus" instead of "physical examination."

Even if the sensitivity and specificity of PSIs were at a theoretical 100%, one also needs to address the conceptual

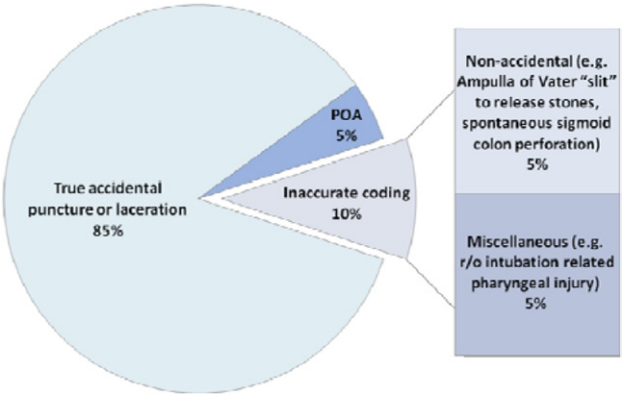


Figure 4. Analysis of false positive cases of accidental punctures or lacerations (APLs). The percentages reported in the figure refer to the percentage of the total number of cases; those reported in the text of the manuscript refer to the percentage of the false positive cases only. POA, present on admission; r/o, rule out.

framework that correlates the occurrence of an APL, iPTX, or pPE/DVT with suboptimal or poor quality of care. In other words, discussion of what actually makes a good quality indicator or measure goes beyond sensitivities and specificities; it implies a risk-adjusted causal relationship between the provision of certain health services and the occurrence of adverse patient outcomes. This discussion is certainly beyond the scope of this article, which is primarily concerned with examining the criterion validity of the PSIs. The interplay of patient characteristics, hospital/physician/clinical team performance, and the specific circumstances surrounding an episode of care in defining outcomes is far from simple (even with risk adjustment), complicating identification of specific adverse events for the purpose of quality benchmarking and public reporting.

Clinical implications

In addition to the policy implications of the results presented here, analyses of the true safety-related events (pPE/DVT, iPTX, and APL) provide us with significant insights into the nature of the adverse events and the circumstances surrounding them.

Postoperative pulmonary embolus and deep vein thrombosis

Orthopaedic and abdominal procedures accounted for more than half the cases in which a postoperative VTE occurred, stressing the importance of DVT mechanical and pharmacologic prophylaxis in noncardiac surgery in general, and particularly, in orthopaedic and abdominal procedures. It is also notable that 17% of the DVTs diagnosed did not involve lower extremity veins, a classical target for quality improvement efforts. The importance of such a finding stems from the fact that the mechanisms and preventive methods involved in upper versus lower extremity DVTs are essentially distinct. Although sequential compression devices, prophylactic anticoagulation, and early ambulation help prevent lower extremity vein thrombosis, timely removal of central venous access lines, such as central venous catheters or peripherally-inserted central catheters (PICC) lines, is more relevant in the prevention of upper extremity DVTs. We believe that risk-adjusted upper and lower extremity DVTs are both important as potential "quality measures;" therefore, modification of this specific PSI and associated ICD-9-CM codes to separately detect lower and upper extremity DVTs would be useful from a quality improvement perspective. The all-cause inpatient mortality of patients who sustained a VTE was elevated (19%), although it is extremely hard to tease out the mortality risk attributable to the VTE alone because many of these patients had multiple serious medical problems and complications during the same admission.

Iatrogenic pneumothorax

Interestingly, only 38% of the iatrogenic pneumothoraces detected could be attributed to central line insertion, and procedures such as cardiac pacemaker or defibrillator placement, transthoracic needle aspiration, and percutaneous liver biopsy accounted for a significant number of additional pneumothoraces. In addition, the mean body mass index of patients who sustained an iPTX was less than 25 kg/m^2, and a large proportion of patients had airway disease, suggesting that a thin patient with COPD or emphysema might be at higher risk for an iPTX than an obese patient, who might present an otherwise technically challenging body habitus for insertion of a central line or a transthoracic needle drainage. Slightly less than half the procedures causing the pneumothoraces were performed by physicians-in-training, which raises the question of the adequacy of supervision of bedside or interventional procedures, especially central line placements.

Accidental puncture or laceration

As expected, APLs were more common in the chest and abdomen and most commonly involved vascular and intra-abdominal organs. In addition, it seems that this indicator detects a high proportion of injuries occurring during spinal operations (eg, durotomy). Similar to iPTX, a significant number of the injuries occurred when a procedure was being performed by a trainee rather than an attending physician. Such a finding is hard to interpret without conducting further research, but raises issues related to the importance of trainee supervision and the learning curve of complex surgical procedures. More importantly, the nature of the accidental punctures and lacerations detected by this indicator presents serious concerns with regard to the discriminatory ability of this indicator. At least 27% of these "true positive" cases revealed injuries that can be considered to have no real clinical (or surgical) relevance. Correlating an injury, such as a serosal bowel tear during an abdominal procedure characterized by extensive intra-abdominal adhesions, to the quality of care that the surgeon, the surgical team, or the hospital provides seems unfair at the least and even ridiculous at times. Adopting such a PSI before improving its ability to distinguish between clinically relevant and clinically irrelevant injuries will provide surgeons with incentives to avoid noting the occurrence of such minor injuries in their operative notes in fear of equating these with serious accidental injuries as indicated by the PSI. Such unintended consequences will clearly be counterproductive to efforts aimed at improving the quality of surgical care.

Limitations

This study is based on retrospective review of medical records of a limited sample of male veteran patients. One might argue that there is less financial incentive in the VA to strive for accuracy in coding in the VA system compared with the private sector, but previous studies show very good accuracy and reliability of diagnostic and procedural coding in the VA.[24,25] In addition, we had limited objective ability to assess the preventability of the detected adverse events or their association with the quality of care provided. Nonetheless, if the study's results are placed in the context of similar validation efforts taking place in several medical institutions across the nation, a reliable idea of the performance of these indicators can be achieved. The utility of these PSIs, particularly pPE/DVT, as performance measures, needs to be viewed with caution and should be reassessed.

Future directions

As they currently stand, the use of some PSIs (particularly pPE/DVT) as quality measures for hospital safety profiling, public reporting, and pay-for performance is premature. Our research team is currently studying the sensitivity of pPE/DVT, iPTX, and APL and the processes of care that might have contributed to the occurrence of these adverse events; this task is necessary for a full evaluation of PSIs, but requires complex epidemiologic statistical models and extensive medical record review to detect safety-related events that are potentially missed by these indicators.

Author Contributions

Study conception and design: Kaafarani, Borzecki, Itani, Rosen

Acquisition of data: Kaafarani, Borzecki, Loveland, Mull, Hickson, MacDonald, Shin, Rosen

Analysis and interpretation of data: Kaafarani, Borzecki, Itani, Loveland, Mull, Shin, Rosen

Drafting of manuscript: Kaafarani, Borzecki, Itani, Loveland, Mull, Hickson, MacDonald, Shin, Rosen

Critical revision: Kaafarani, Borzecki, Itani, Rosen

REFERENCES

1. Institute of Medicine. To Err Is Human: Building a Safer Health System. Washington, DC: National Academy; 2000.
2. Institute of Medicine. Crossing the Quality Chasm. Washington, DC: National Academy; 2001.
3. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA 2003;290:1868–1874.
4. Miller DC, Filson CP, Wallner LP, et al. Comparing performance of Morbidity and Mortality Conference and National Surgical Quality Improvement Program for detection of complications after urologic surgery. Urology 2006;68:931–937.
5. Hutter MM, Rowell KS, Devaney LA, et al. Identification of surgical complications and deaths: an assessment of the traditional surgical morbidity and mortality conference compared with the American College of Surgeons-National Surgical Quality Improvement Program. J Am Coll Surg 2006;203:618–624.
6. Romano PS, Geppert JJ, Davies S, et al. A national profile of patient safety in U.S. hospitals. Health Aff (Millwood) 2003;22:154–166.
7. AHRQ Quality Indicators: Guide to Patient Safety Indicators. Version 3.0a. Rockville, MD: Agency for Healthcare Research and Quality (AHRQ); March 2003; version 3.1.
8. Healthgrades patient safety methodology. Available at http://www.healthgrades.com/media/DMS/pdf/PatientSafetyMethodology.pdf. Accessed April 15, 2010.
9. National Quality Forum. NQF endorsed standards. Available at: http://www.qualityforum.org/Measures_List.aspx. Accessed April 15th, 2010.
10. Centers for Medicare and Medicaid Services. Details for: Quality measures for reporting in fiscal year 2009 for 2010 update. Available at: http://www.cms.hhs.gov/apps/media/press/factsheet.asp?Counter=3044&intNumPerPage=10&checkDate=&checkKey=&srchType=1&numDays=3500&srchOpt=0&srchData=&keywordType=All&chkNewsType=6&intPage=&showAll=&pYear=&year=&desc=false&cboOrder=date. Accessed April 15, 2010.
11. QualityNet. Available at: http://www.qualitynet.org. Accessed April 15, 2010.
12. Bahl V, Thompson MA, Kau TY, et al. Do the AHRQ patient safety indicators flag conditions that are present at the time of hospital admission? Med Care 2008;46:516–522.
13. Houchens RL, Elixhauser A, Romano PS. How often are potential patient safety events present on admission? Jt Comm J Qual Patient Saf 2008;34:154–163.
14. White RH, Sadeghi B, Tancredi DJ, et al. How valid is the ICD-9-CM based AHRQ patient safety indicator for postoperative venous thromboembolism? Med Care 2009;47:1237–1243.
15. Henderson KE, Recktenwald A, Reichley RM, et al. Clinical validation of the AHRQ postoperative venous thromboembolism patient safety indicator. Jt Comm J Qual Patient Saf 2009;35:370–376.
16. Utter GH, Zrelak PA, Baron R, et al. Positive predictive value of the AHRQ accidental puncture or laceration patient safety indicator. Ann Surg 2009;250:1041–1045.
17. Weller WE, Gallagher BK, Cen L, Hannan EL. Readmissions for venous thromboembolism: expanding the definition of patient safety indicators. Jt Comm J Qual Saf 2004;30:497–504.
18. Gallagher B, Cen L, Hannan EL. Validation of AHRQ's patient safety indicator for accidental puncture or laceration. Available at: http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=aps2&part=A1586. Accessed April 15, 2010.
19. Zhan C, Battles J, Chiang YP, Hunt D. The validity of ICD-9-CM codes in identifying postoperative deep vein thrombosis and pulmonary embolism. Jt Comm J Qual Patient Saf 2007;33:326–331.

20. Kaafarani HM, Rosen AK. Using administrative data to identify surgical adverse events: an introduction to the Patient Safety Indicators. Am J Surg 2009;198:S63–68.
21. Quality AfHRa. Patient Safety Indicators: technical specifications, March 2003, version 3.1.
22. Sadeghi B, Baron R, Zrelak P, et al. Cases of iatrogenic pneumothorax can be identified from ICD-9-CM coded data. Am J Med Qual 2010;25:218–224.
23. Allison JJ, Wall TC, Spettell CM, et al. The art and science of chart review. Jt Comm J Qual Improv 2000;26:115–136.
24. Kashner TM. Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. Med Care 1998;36:1324–1336.
25. Petersen LA, Wright S, Normand SL, Daley J. Positive predictive value of the diagnosis of acute myocardial infarction in an administrative database. J Gen Intern Med 1999;14:555–558.

AIDS Care
Vol. 23, No. 8, August 2011, 947–956

Routledge
Taylor & Francis Group

Recent cigarette smoking and HIV disease progression: no evidence of an association

Conrad Kabali[a,b]*, Debbie M. Cheng[c,d], Daniel R. Brooks[a], Carly Bridden[d], C. Robert Horsburgh Jr.[a] and Jeffrey H. Samet[d,e]

[a]*Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA;* [b]*National Institute of Medical Research, Dar es Salaam, Tanzania;* [c]*Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA;* [d]*Department of Medicine, Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA;* [e]*Department of Community Health Sciences, Boston University School of Public Health, Boston, MA, USA*

(*Received 27 April 2010; final version received 18 November 2010*)

The association between smoking and HIV disease progression has been examined in several studies; however, findings have been inconsistent. We examined the effect of recent cigarette smoking on CD4$^+$ T cell count/μl (CD4 count) and HIV RNA concentration (HIV viral load (VL)) among two HIV-infected cohorts with alcohol problems in Massachusetts in the periods 1997–2001 and 2001–2006 using a prospective cohort design and linear mixed models. Smoking groups were defined as: minimal or non-smokers, light smokers, moderate smokers, and heavy smokers. Age, alcohol use, injection drug use, depressive symptoms, gender, annual income, and antiretroviral therapy adherence were considered as potential confounders. Among 462 subjects, no significant differences in CD4 count or VL were found between smoking groups. Using minimal or non-smokers as the reference group, the adjusted mean differences in CD4 count were: 8.2 (95% confidence interval (CI): −17.4, 33.8) for heavy smokers; −0.1 (95% CI: −25.4, 5.1) for moderate smokers; and −2.6 (95% CI: −28.3, 3.0) for light smokers. For log10 VL, the adjusted differences were: 0.03 (95% CI: −0.12, 0.17) for heavy smokers; −0.06 (95% CI: −0.20, 0.08) for moderate smokers; and 0.14 (95% CI −0.01, 0.28) for light smokers. This study did not find an association between smoking cigarettes and HIV disease progression as measured by CD4 cell count and VL.

Keywords: cigarette smoking; CD4$^+$ T cells; viral load; HIV

Introduction

With the major prognostic advance of highly active antiretroviral therapy (HAART; Crum et al., 2006; Detels et al., 1998; Palella et al., 1998), impetus to understand other potential avenues to prevent disease progression among Human Immunodeficiency Virus (HIV)-infected persons has been sought (Baum et al., 1995; Cheng et al., 2007; Cook et al., 2008; Fawzi et al., 2004; Jia et al., 2007). Since smoking is common among HIV-infected persons (Niaura et al., 2000; Webb, Vanable, Carey, & Blair, 2007) its effect on HIV disease progression has merited study. Smoking can suppress the maturation of dendritic cells in the lymph nodes thereby weakening the function of CD4$^+$ T cells (Robbins et al., 2004; Robbins, Franco, Mouded, Cernadas, & Shapiro, 2008). It can also affect the efficiency of peripheral blood mononuclear cells to secrete cytokines (Ouyang et al., 2000). Smoking has also been reported to up-regulate the expression of Fas (cell surface molecules mediating apoptotic cell death) on peripheral blood lymphocytes, rendering them susceptible to apoptosis (Bijl, Horst, Limburg, & Kallenberg, 2001). Other reported possible mechanisms explaining the adverse effects of smoking on immunological function have been described (Abbud, Finegan, Guay, & Rich, 1995; Carrillo, Castro, Cuevas, Diaz, & Cabrera, 1991; Kalra, Singh, Savage, Finch, & Sopori, 2000; Kuniak et al., 1995; Petersen, Steimel, & Callaghan, 1983; Silverman, Potvin, Alexander Jr, & Chretien, 1975; Sopori, Gairola, DeLucia, Bryant, & Cherian, 1985; Sopori & Kozak, 1998; Tollerud et al., 1989a).

Several epidemiologic studies have investigated the relation between cigarette smoking and the course of HIV infection yielding mixed results. Two cohort studies (Conley et al., 1996; Crothers et al., 2005) and two cross-sectional studies (Palacio, Hilton, Canchola, & Greenspan, 1997; Slavinsky III et al., 2002) found an association between smoking and development of opportunistic infections (OIs). In contrast, eight cohort studies (Burns et al., 1996; Coates et al., 1990; Craib et al., 1992; Eskild & Petersen, 1994; Galai et al., 1997; Nieman, Fleming, Coker, William, & Mitchell, 1993; Stephenson et al., 1999; Webber, Schoenbaum, Gourevitch, Buono, & Klein, 1999) and two cross-sectional studies (Gritz, Vidrine, Lavez,

*Corresponding author. Email: ckabali@bu.edu

ISSN 0954-0121 print/ISSN 1360-0451 online

DOI: 10.1080/09540121.2010.542128
http://www.informaworld.com

Amick, & Arduino, 2004; Webb et al., 2007) reported a null association. One study (Royce & Winkelstein, 1990) linked smoking with an increase in $CD4^+$ T cells among males although the increase was less pronounced in HIV-infected individuals.

Inconsistent results regarding the effect of smoking on HIV disease progression may in part be attributed to the differing characteristics of the study populations. Adjustment for potential confounders (e.g., alcohol use) did not consistently occur. As most previous epidemiologic studies used the incidence of OIs as a primary endpoint, we considered the other useful biological markers that might complement the observations of clinical outcomes in the examination of whether smoking accelerates HIV disease progression.

We therefore analyzed data from a prospectively assessed two cohorts of HIV-infected patients with alcohol problems to examine the association of cigarette smoking with $CD4^+$ T cell count (CD4 count) and HIV viral load (VL), We hypothesized that smoking would be associated with a lower CD4 count and a higher VL.

Design and methods

Study population

Study participants were from two longitudinal cohorts of HIV-infected persons with alcohol problems: HIV Alcohol Longitudinal Cohort (HIV-ALC) study only (1997–2001; $n = 78$); HIV Longitudinal Interrelationships of Viruses and Ethanol (HIV-LIVE) study only (2001–2006; $n = 230$); and both HIV-ALC and HIV-LIVE studies ($n = 154$). Eligibility criteria and recruitment methods for HIV-ALC (Samet, Horton, Traphagen, Lyon, & Freedberg, 2003) and HIV-LIVE (Samet et al., 2007) were the same with follow-up visits planned every six months. Inclusion criteria included a documented HIV antibody test, a history of alcohol problems as measured by the CAGE questionnaire or a clinical investigator's assessment (Samet, Phillips, & Horton, 2004), age 18 years or older, ability to speak English or Spanish, and at least one contact person to assist with the follow-up. Exclusion criteria included score <21 on the 30-item Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975; Smith, Horton, Saitz, & Samet, 2006), inability to provide informed consent or answer the interview questions, and plans to move from the Boston area in the subsequent 12 months. For the current analysis, we also excluded those who did not have at least one follow-up visit. Laboratory measurements were obtained at each interview. Additional details on this population have been provided elsewhere (Samet et al., 1995, 2007). The study was approved by the Boston Medical Center and Beth Israel Deaconess Medical Center Institutional Review Boards.

Outcome assessment

The primary outcomes for this analysis were $CD4^+$ T cell count/μl (CD4 count) and log10 plasma HIV RNA/ ml (VL). CD4 count was determined by flow cytometry at the hospital laboratories. VL was measured using a branched-chain assay (lowest detection threshold = 75 copies/ml) or a polymerase chain reaction (lowest detection threshold = 50 copies/ml for ultrasensitive assay, and 500 copies/ml for standard assay; Pachl et al., 1995).

Exposure assessment

Information on smoking was collected at each study visit. Smoking status was categorized per Okuyemi et al. (2002, 2004) as follows: Minimal or non-smoker (smoked less than one cigarette per day); light smoker (smoked one to less than 10 cigarettes per day); moderate smoker (smoked 10 to less than 20 cigarettes per day); or heavy smoker (smoked 20 or more cigarettes per day). We used non-parametric local linear polynomial curves (*loess*) (Cleveland, Grosse, & Shyu, 1992) to verify that the cutoffs were reasonable for our data. We also performed a secondary, confirmatory analysis including smoking categorized based on quintiles of the distribution.

We defined a smoker as someone who answered affirmatively to the question "Do you currently smoke cigarettes?" As a secondary analysis, we defined smoking based on the subject's reported smoking status at two successive visits. That is, the outcome at each time point was modeled as a function of smoking from the current and previous study visit, thus accounting for whether the subjects changed or maintained their recent smoking behavior. In the case of missed visits, smoking status from the last available visit was used. We categorized smoking as follows: consistent smokers (smoked in two consecutive visits); consistent minimal or non-smokers (smoked <1 cigarette per day in two consecutive visits); recent quitters (smoked in the last visit but stopped smoking in the current visit); and new/relapsed smokers (not smoked in the last visit but started or resumed smoking in the current visit). The categorization was based on data from previous studies which suggest that the effect of smoking on the immune system is acute (Hersey, Prendergast, & Edwards, 1983; Sunyer et al., 1996; Tollerud et al., 1989b), with induction period of

about five weeks (Thomas, Holt, & Keast, 1975) to 10 weeks (Chalmer, Holt, & Keast, 1975), and lasts for about 6–35 weeks since quitting (Miller, Goldstein, Murphy, & Ginns, 1982; Radloff, 1977; Thomas, Holt, & Keast, 1975). Since the time elapsed between the last HIV-ALC visit to the first HIV-LIVE visit was too long (range 1–66 months) for some of the subjects enrolled in both cohorts, this secondary analysis was restricted to subjects who participated in HIV-LIVE.

Statistical analysis

We performed descriptive analyses to characterize the study population, overall and by baseline smoking status.

We applied linear mixed effects models to account for correlated measures within subjects. The models included subject-specific random intercepts and slopes, and adjusted for the value of the outcome (i.e., CD4 count and VL) at the previous study visit. We fitted a separate model for each outcome. Graphs illustrating trajectories of HIV disease progression over time were also plotted using outcome estimates from linear mixed effects models.

Age (modeled as a continuous variable), alcohol use (≤ 2 drinks per day, $>2-\leq 4$ drinks per day, and >4 drinks per day), current injection drug use (user vs. non-user), depressive symptoms (CESD score >23 vs. ≤ 23), gender, annual income ($>$median (US$7500) vs. $\leq$ median), and antiretroviral therapy (ART) adherence (not on medication, on medication and adherent, and on medication but not adherent) were considered as potential confounders during analysis. The categories for alcohol use were made narrower than those suggested by Cook et al. (2009) so as to account for residual confounding. The cut-off point for depressive symptoms was determined with Radloff (1977) criteria. ART adherence was measured using the AIDS Clinical Trials Group criteria (Chesney et al., 2000) and defined as a self-report of 100% adherent in the past three days (Samet, Horton, Meli, Freedberg, & Palepu, 2004). Preliminary models were fitted separately for each potential confounder. Confounders were then added sequentially in the models according to the magnitude of their effect on the association between cigarette smoking and outcome. Any potential confounder that changed the point estimate of cigarette smoking by more than 10% was included in the final model. Smoking status and all covariates with the exception of gender and age were analyzed as time-varying variables and updated at each time point. An interaction between smoking and time was included in the models and evaluated for its statistical significance. To assess whether ART modifies the effect of smoking, subgroup analyses were repeated separately for subjects not on ART and for subjects on ART and adherent to medication. Subjects who changed their ART status contributed observations to each stratum depending on the ART status of the observation.

For the primary analysis, undetectable VL was imputed as half the value of the lowest threshold of assay sensitivity. Secondary analyses evaluated the potential bias due to imputing assay values using half the limit of detection (Greenland & Lash, 2008) as follows: The undetectable assay measurements were imputed with plausible values from a logit-logistic distribution (Lesaffre, Rizopoulos, & Tsonaka, 2007) with a scale parameter of 0.8, accounting for the fact that VL values are bounded by 0. The distribution also took into account that 0 VL is unlikely as no current treatment can completely eliminate HIV. This process was repeated in 10,000 simulation runs and 95% simulation intervals (SI) were obtained. Statistical analysis was done using SAS version 9.1 (SAS Institute, Inc., Cary, North Carolina).

Results

The cohort ($n=462$) is described in Table 1 with the following demographic characteristics: black (43%); male (77%); median age 42 years (range 21–71 years); and median annual income of US$7500 or less. In the combined cohort, 77% (358/462) of subjects reported having smoked cigarettes within the last month before enrollment; 23% were minimal or non-smokers, 25% were light, 22% were moderate, and 31% were heavy smokers. The median baseline CD4 count and VL were 380 cells/μl and 1175 copies/ml, respectively. The median follow-up time was 18 months for those enrolled in only HIV-LIVE, 14 months for those in only HIV-ALC, and 40 months for those in both studies. In the overall cohort, the median number of visits was seven (range: 2–14 visits). Subjects who completed the majority of study of visits had higher mean baseline CD4 count (difference = 117.3; $p<0.0001$) and lower mean baseline log10 VL (difference = 0.35; $p=0.03$) than those who did not complete the majority of study visits. There was no significant association between baseline smoking status and whether the subject completed the majority of study visits in this cohort ($\chi^2=0.41$; $p=0.52$). Subjects contributed 3141 observations across all follow-up visits: 827 (26%) classified as minimal or non-smokers, 696 (22%) light smokers, 772 (25%) moderate smokers, and 846 (27%) as heavy smokers.

Table 1. Sociodemographic and clinical characteristics of HIV-infected persons with a history of alcohol problems in two prospective cohorts stratified by baseline smoking status ($n = 462$).

	n (%)					
Covariates	Total *n* (%) $n = 462$	Minimal or non-smokers $n = 104$	Light smokers $n = 115$	Mod. Smokers $n = 100$	Heavy smokers $n = 143$	Mean no. cigarretes[a] smoked per day (SD) $n = 358$
Sociodemographic variables						
Age						
≤30	22 (5)	5 (5)	6 (5)	3 (3)	8 (6)	14.18 (8.79)
31–40	174 (38)	34 (33)	47 (41)	43 (43)	50 (35)	14.11 (10.14)
>40	266 (58)	65 (63)	62 (54)	54 (54)	85 (59)	13.79 (9.00)
Gender						
Females	108 (23)	81 (78)	86 (75)	73 (73)	114 (80)	14.30 (9.59)
Males	354 (77)	23 (22)	29 (25)	27 (27)	29 (20)	12.72 (8.87)
Race						
Black	198 (43)	43 (41)	64 (56)	57 (57)	34 (24)	10.92 (7.08)
White	154 (33)	38 (37)	15 (13)	21 (21)	80 (56)	19.28 (10.85)
Hispanic	87 (19)	18 (17)	29 (25)	17 (17)	23 (16)	12.18 (8.43)
Other	23 (5)	5 (5)	7 (6)	5 (5)	6 (4)	14.36 (7.15)
Income						
>Median (US$7500)	215 (47)	55 (53)	43 (37)	48 (48)	69 (49)	15.60 (10.02)
< = Median (US$7500)	245 (53)	49 (47)	72 (63)	51 (52)	73 (51)	12.63 (8.79)
Average no. drinks/day						
0–2	381 (83)	91 (88)	93 (81)	80 (81)	117 (82)	13.87 (9.42)
>2–4	30 (7)	8 (8)	11 (10)	6 (6)	5 (4)	12.86 (8.16)
>4	50 (11)	5 (5)	11 (10)	13 (13)	21 (15)	14.96 (10.13)
Injecting drug use						
User	79 (17)	7 (7)	14 (12)	25 (25)	33 (23)	15.96 (9.43)
Non user	383 (83)	97 (93)	101 (88)	75 (75)	110 (77)	13.60 (9.41)
ART status						
Not on meds	173 (38)	35 (34)	34 (30)	41 (41)	63 (44)	13.98 (9.04)
On meds, not adherent	84 (18)	14 (13)	31 (27)	13 (13)	26 (18)	15.35 (11.61)
On meds, adherent	204 (44)	55 (53)	50 (43)	45 (45)	54 (38)	13.42 (8.90)
Depressive symptoms						
Depressed	217 (47)	35 (34)	55 (47)	48 (48)	79 (55)	15.08 (9.95)
Not Depressed	245 (53)	69 (66)	60 (52)	52 (52)	64 (45)	12.92 (8.86)
Cohort						
ALC-only	78 (17)	17 (16)	15 (13)	18 (18)	28 (20)	16.82 (13.30)
LIVE-only	230 (50)	58 (56)	52 (45)	52 (52)	68 (48)	13.91 (8.16)
Combined	154 (33)	29 (28)	48 (42)	30 (30)	47 (33)	13.46 (9.64)
Median baseline CD4 count	390	383	410	324	422	
Median baseline viral load	1188	388	1723	1382	1251	

[a]Minimal or non-smokers are excluded.

The plots in Figure 1 show the unadjusted mean CD4 count or VL over time across smoking categories. The plots suggest potential variation in mean differences in CD4 count or VL over time between smoking categories. However, the smoking-time interaction term was not statistically significant in any

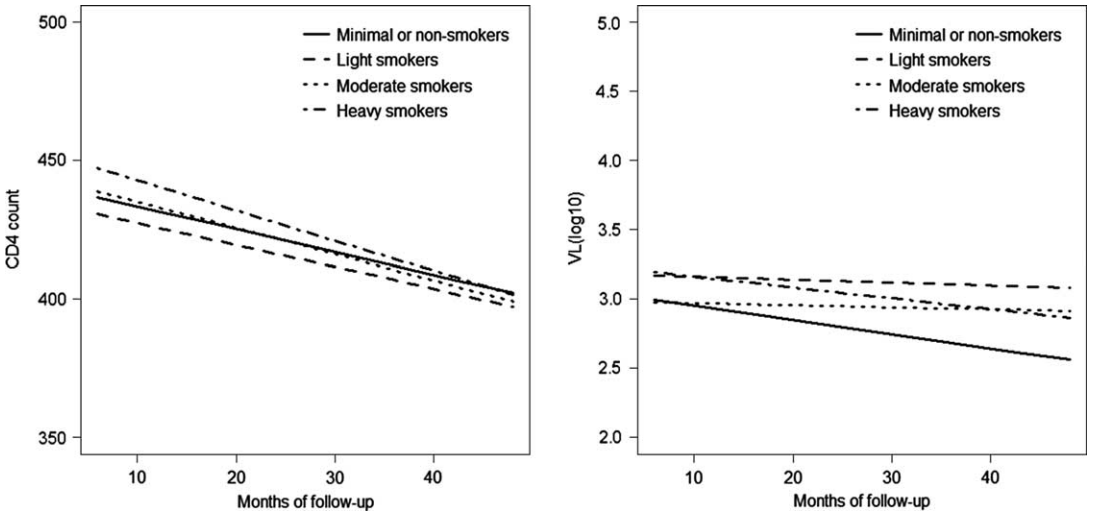


Figure 1. Unadjusted mean CD4 count and VL over time.

of the regression models and was therefore excluded from subsequent analyses.

We did not find any substantial differences in CD4 count or VL across categories of smoking (Table 2). Using minimal or non-smokers as the reference group, the adjusted mean differences in CD4 count were: 8.2 (95% confidence interval (CI): −17.4, 33.8; $p = 0.44$) for heavy smokers; −0.1 (95% CI: −25.4, 5.1; $p = 0.48$) for moderate smokers; and −2.6 (95% CI: −28.3, 3.0; $p = 0.90$) for light smokers. For log10 VL, the adjusted differences were: 0.03 (95% CI: −0.12, 0.17; $p = 0.39$) for heavy

Table 2. The association between smoking status and markers of HIV disease progression.

	Mean differences in CD4 count[a]		Mean differences in HVL[b]	
	All subjects combined ($n = 462$)[c]			
Smoking status	Crude [95% CI]	Adjusted [95% CI]	Crude [95% CI]	Adjusted [95% CI]
Heavy smokers	−3.2 [−27.7, 21.2]	8.2 [−17.4, 33.8]	0.18 [0.03, 0.32]	0.03 [−0.12, 0.17]
Moderate smokers	−11.7 [−36.0, 12.6]	−0.1 [−25.4, 25.1]	0.11 [−0.04, 0.25]	−0.06 [−0.20, 0.08]
Light smokers	−13.0 [−37.8, 11.9]	−2.6 [−28.3, 23.0]	0.27 [0.12, 0.42]	0.14 [−0.01, 0.28]
Minimal or non smokers	Reference	Reference	Reference	Reference
		Not on ART[d,e]		
Heavy smokers	24.0 [−22.4, 70.3]	27.4 [−19.7, 74.4]	−0.11 [−0.36, 0.14]	−0.09 [−0.35, 0.17]
Moderate smokers	27.7 [−18.0, 73.4]	30.1 [−16.0, 76.1]	−0.22 [−0.47, 0.03]	−0.20 [−0.56, 0.06]
Light smokers	13.8 [−30.7, 58.4]	16.0 [−28.9, 60.9]	−0.15 [−0.39, 0.09]	−0.14 [−0.38, 0.11]
Minimal or non-smokers	Reference	Reference	Reference	Reference
		On ART and adhered to medication[d,f]		
Heavy smokers	−8.4 [−34.3, 17.5]	−6.4 [−32.4, 19.6]	0.05 [−0.14, 0.24]	0.03 [−0.16, 0.21]
Moderate smokers	−14.5 [−40.4, 11.3]	−14.2 [−40.0, 11.6]	0.05 [−0.14, 0.23]	0.01 [−0.18, 0.19]
Light smokers	−11.2 [−38.2, 15.8]	−11.4 [−38.3, 15.6]	0.25 [0.06, 0.45][g]	0.24 [0.04, 0.44][g]
Minimal or non smokers	Reference	Reference	Reference	Reference

[a]Adjusted analyses controlled for previous CD4$^+$ cell count, ART status, time, and depressive symptoms.
[b]Adjusted analyses controlled for previous log10 HIV RNA, ART status, income, depressive symptoms, injection drug use, age, alcohol, time, and gender.
[c]No. observations was 3141.
[d]ART was removed in the model because we stratified on it.
[e]No. subjects was 291. No. observations was 1026. Included only observations when subject was on ART.
[f]No. subjects was 363. No. observations was 1586. Included only observations when subject was on ART and adhered to medication.
[g]Results are statistically significant.

smokers; −0.06 (95% CI: −0.20, 0.08; $p = 0.83$) for moderate smokers; and 0.14 (95% CI: −0.01, 0.28; $p = 0.06$) for light smokers.

When analysis was stratified by ART status, no clear evidence of an association between smoking with CD4 count or HIV VL was found (Table 2). We observed a small but statistically significant increase in VL for light smokers compared with minimal or non-smokers among subjects who adhered to ART (the adjusted mean difference was 0.24; 95% CI: 0.04, 0.44; $p = 0.01$). However, similar associations were not observed for categories of heavier smoking.

Similarly, no substantial outcome differences were observed for subjects who switched or maintained their smoking behavior (Table 3).

Results (not shown) remained similar in analyses examining the potential bias due to the imputed assay measurements, suggesting that the unobserved values in the lower range did not have a large impact on results. Similarly, the analysis with smoking categorized based on quintiles did not alter the study findings.

Discussion

This study does not provide evidence that cigarette smoking is associated with a decrease in CD4 count or an increase in VL among HIV-infected patients. Specifically, we observed no substantial differences in CD4 count or VL between minimal or non-smokers, light smokers, moderate smokers, and heavy smokers. Moreover, we found no substantial differences in CD4 count or VL when smoking status changed in two consecutive visits.

Our findings are in accordance with previous cohort studies which did not detect an association between cigarette smoking and HIV disease progression (Burns et al., 1996; Coates et al., 1990; Craib et al., 1992; Eskild & Petersen, 1994; Galai et al., 1997; Stephenson et al., 1999; Webber et al., 1999) when using the onset of an AIDS defining condition as a primary outcome. Our design, using CD4 count and VL as markers for HIV disease progression, is well suited for HIV cohorts with less advanced disease, fewer anticipated OIs.

Results from this study are in contrast with findings from two cohort studies which reported that smoking enhances HIV disease progression using OIs as outcomes (Conley, Bush, Buchbinder, & Penley, 1996; Crothers et al., 2005). In one of these studies CD4 counts were also examined (Conley et al., 1996), but not associated with smoking. As smoking may selectively affect target organs (e.g., lungs), the latter study underscores the need to investigate biological markers of immunological dysfunction in addition to OIs in order to assess its impact on the immune system.

Although, we found no evidence of a relation between smoking and HIV disease progression, we could not rule out the potential effect of smoking on CD4 cells or HIV which are not in peripheral blood. As high concentrations of smoke can get trapped in the lungs, it would not be surprising if most of the affected cells are those surrounding the lungs. For example, one study (Wewers et al., 1998) found that, compared to minimal or non-smokers, HIV-infected smokers had a significant depletion in CD4 cells in their *bronchoalveolar lavage* fluid. Analyzing samples of immunological markers derived from sites other than the peripheral blood may shed some light on the issue.

Our study was conducted using a cohort of patients with current or past alcohol problems. As alcohol drinkers tend to smoke more than non-alcohol drinkers (Collins & Marks, 1995), the choice of this cohort ensured availability of ample smokers for analysis. The selection of this cohort came with a caution in that the generalizability of its findings may be limited to a population with alcohol problems. However, given the prevalence of past alcohol

Table 3. The association between smoking status at two successive visits and markers of HIV disease progression ($n = 383$).[a]

	Mean difference in CD4 count [95% CI]		Mean difference in HVL [95% CI]	
	Crude	Adjusted[b]	Crude	Adjusted[c]
Smoking status				
New/relapsed smokers	16.87 [−6.60, 60.35]	22.73 [−0.28, 65.74]	−0.11 [−0.41, 0.19]	−0.11 [−0.39, 0.17]
Consistent smokers	−12.14 [−30.00, 5.69]	−2.25 [−20.18, 15.68]	0.18 [0.05, 0.30]	0.11 [−0.05, 0.26]
Recent quitters	−32.39 [−71.00, 6.23]	−33.20 [−71.39, 4.98]	−0.07 [−0.34, 0.20]	0.04 [−0.22, 0.29]
Consistent minimal or non-smokers	Reference	Reference	Reference	Reference

[a]Analysis restricted to HIV LIVE cohort. No. observations was 1974. All p values > 0.05.
[b]Adjusted for CD4 count at previous visit, ART status, age, time, and alcohol use.
[c]Adjusted for HVL at previous visit, ART status, time, and depressive symptoms.

problems among HIV-infected individuals, up to 40% (Samet, Phillips, & Horton, 2004), these findings would still be of importance even if not applicable to a non-alcohol affected HIV population. An interaction term between smoking and current alcohol drinking (categorized by dose) was not statistically significant in regression models.

The main strength of this study was the inclusion of assessments of changes in immunological biological markers over time using repeated measures on the same individuals. We identified the use of such methodology in only one other study (Sunyer et al., 1996) among HIV seronegative subjects in which a positive association existed between smoking and an increase in white blood cells. We performed post-hoc power calculations to assess the differences in CD4 count our study could detect with reasonably high power. While we utilized longitudinal regression methods in the analyses, for the purposes of power calculations we considered a simpler setting utilizing a single time point. Thus, our estimates are conservative as the longitudinal analyses are expected to increase the study power. Assuming a standard deviation of 42 (based on our observed data at baseline), the minimum detectable difference in mean CD4 count between any two smoking groups that our sample size could detect with 80% power was 16 cells/μl. hence this study was adequately powered to detect effect sizes observed previously (Sunyer et al., 1996), a difference in CD4 count of 74 cells/μl between heavy smokers and never smokers. Other strengths included the ability to analyze short term effects of smoking initiation and cessation, information on smoking dosage, and the availability of many important potential confounders.

We note our study had limitations. First, information on cigarette smoking was self-reported. Second, the number of cigarettes smoked may not necessarily reflect the amount of smoke inhaled, underestimating the actual smoking dosage.

In summary, we did not find significant associations between cigarette smoking and CD4 count or VL among HIV-infected patients with alcohol problems. Future epidemiologic studies concerning the impact of smoking on HIV disease progression may provide more insight if focused on this substance's effect on specific tissues and particular immunological systems in addition to immunological markers (i.e., CD4 count and VL) in the peripheral blood.

Acknowledgements

This study was supported by funding from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) RO1-AA11785, RO1-AA13216, K24-AA015674 and Fogarty International Center D43-TW006808. The authors thank Emily Quinn for her data management support. We also appreciate the collaboration of Dr Howard Libman at Beth Israel Deaconess Medical Center who led the recruitment of participants at that institution.

References

Abbud, R.A., Finegan, C.K., Guay, L.A., & Rich, E.A. (1995). Enhanced production of human immunodeficiency virus type 1 by in vitro-infected alveolar macrophages from otherwise healthy cigarette smokers. *The Journal of Infectious Diseases*, *172*, 859–863.

Baum, M.K., Shor-Posner, G., Lu, Y., Rosner, B., Sauberlich, H.E., Fletcher, M.N., & Hennekens, C.H. (1995). Micronutrients and HIV-1 disease progression. *AIDS*, *9*, 1051–1056.

Bijl, M., Horst, G., Limburg, P.C., & Kallenberg, C.G. (2001). Effects of smoking on activation markers, Fas expression and apoptosis of peripheral blood lymphocytes. *European Journal of Clinical Investigation*, *31*, 550–553.

Burns, D.N., Hillman, D., Neaton, J.D., Sherer, R., Mitchell, T., Capps, L., & Michael, D. (1996). Cigarette smoking, bacterial pneumonia, and other clinical outcomes in HIV-1 infection. *Journal of Acquired Immune Deficiency Syndrome*, *13*, 374–383.

Carrillo, T., Castro, F.R., Cuevas, M., Diaz, F., & Cabrera, P. (1991). Effect of cigarette smoking on the humoral immune response in pigeon fanciers. *Allergy*, *46*, 241–244.

Chalmer, J., Holt, P., & Keast, D. (1975). Cell-mediated responses to transplanted tumors in mice chronically exposed to cigarette smoke. *Journal of the National Cancer Institute*, *55*, 1129–1134.

Cheng, D.M., Nunes, D., Libman, H., Vidaver, J., Alperen, J.K., Saitz, R., & Samet, J. (2007). Impact of hepatitis C on HIV progression in adults with alcohol problems. *Alcoholism: Clinical and Experimental Research*, *31*, 829–836.

Chesney, M.A., Ickovics, J.R., Chambers, D.B., Gifford, A.L., Neidig, J., Zwickl, B., & Wu, A.W. (2000). Patient care committee and adherence working group of the outcomes committee of the Adult AIDS Clinical Trials Group (AACTG). Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: The AACTG adherence instrument. *AIDS Care*, *12*, 255–266.

Cleveland, W.S., Grosse, E., & Shyu, W.M. (1992). Local regression models. In J.M. Chambers & T.J. Hastie (Eds.), *Statistical models in S* (pp. 309–373). New York: Chapman & Hall/CRC.

Coates, R.A., Farewell, V.T., Raboud, J., Read, S.E., Macfadden, D.K., Calzavara, L.M., & Fanning, M.M. (1990). Cofactors of progression to acquired immunodeficiency syndrome in a cohort of male sexual contacts of men with human immunodeficiency virus disease. *American Journal of Epidemiology*, *132*, 717–722.

Collins, A.C., & Marks, M.J. (1995). Animal models of alcohol-nicotine interactions. In J.B. Fertig & J.P. Allen (Eds.), *Alcohol and tobacco: From basic science to clinical practice*. NIAAA Research Monograph No. 30. (pp. 129–143). NIH Pub. 95–3931.

Conley, L.J., Bush, T.J., Buchbinder, S.P., & Penley, K.A. (1996). The association between cigarette smoking and selected HIV-related medical conditions. *AIDS*, *10*, 1121–1126.

Cook, J.A., Burke-Miller, J.K., Cohen, M.H., Cook, R.L., Vlahov, D., Wilson, T.E., & Grey, D.D. (2008). Crack cocaine, disease progression, and mortality in a multicenter cohort of HIV-1 positive women. *AIDS*, *22*, 355–1363.

Cook, R.L., Zhu, F., Belnap, B.H., Weber, K., Cook, J.A., Vlahov, D., … Cohen, M.H. (2009). Longitudinal trends in hazardous alcohol consumption among women with human immunodeficiency virus infection, 1995–2006. *American Journal of Epidemiology*, *169*, 1025–1032.

Craib, K.J., Schechter, M.T., Montaner, J.S.G., Le, T.N., Sestak, P., Willoughby, B., & O'Shaughnessy, M.V. (1992). The effect of cigarette smoking on lymphocyte subsets and progression to AIDS in a cohort of homosexual men. *Clinical and Investigative Medicine*, *15*, 301–308.

Crothers, K., Griffith, T.A., McGinnis, K.A., Rodriguez-Barradas, M.C., Leaf, D.A., Weissman, S., & Justice, A.C. (2005). The impact of cigarette smoking on mortality, quality of life, and comorbid illness among HIV-positive veterans. *Journal of General Internal Medicine*, *20*, 1142–1145.

Crum, N.F., Riffenburgh, R.H., Wegner, S., Agan, B.K., Tasker, S.A., Spooner, K.M., & Mark, R. (2006). Comparisons of causes of death and mortality rates among HIV-infected persons: Analysis of the pre-, early, and late HAART (Highly Active Antiretroviral Therapy) eras. *Journal of Acquired Immune Deficiency Syndromes*, *41*, 194–200.

Detels, R., Munoz, A., McFarlane, G., Kingsley, L.A., Margolick, J.B., Giorgi, J., & Phair, J.P. (1998). Effectiveness of potent antiretroviral therapy on time to AIDS and death in men with known HIV infection duration. Multicenter AIDS Cohort Study. *The Journal of the American Medical Association*, *280*, 1497–1503.

Eskild, A., & Petersen, G. (1994). Cigarette smoking and drinking alcohol are not associated with rapid progression to acquired immunodeficiency syndrome among homosexual men in Norway. *Scandinavian Journal of Social Medicine*, *3*, 209–212.

Fawzi, W.W., Msamanga, G.I., Spiegelman, D., Wei, R., Kapiga, S., Villamor, E., & Hunter, D.J. (2004). A randomized trial of multivitamin supplements and HIV disease progression and mortality. *The New England Journal of Medicine*, *351*, 23–32.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Galai, N., Park, L.P., Wesch, J., Visscher, B., Riddler, S., & Margolick, J.B. (1997). Effect of smoking on the clinical progression of HIV-1 infection. *Journal of Acquired Immune Deficiency Syndromes*, *14*, 451–458.

Greenland, S., & Lash, T.L. (2008). Bias analysis. In K.J. Rothman, S. Greenland, & T.L. Lash (Eds.), *Modern epidemiology* (3rd ed., pp. 345–380). Philadelphia, PA: Lippincott Williams & Wilkins.

Gritz, E.R., Vidrine, D.J., Lavez, A.B., Amick, B.C., & Arduino, R.C. (2004). Smoking behavior in a low-income multiethnic HIV/AIDS population. *Nicotine & Tobacco Research*, *6*, 71–77.

Hersey, P., Prendergast, D., & Edwards, A. (1983). Effects of cigarette smoking on the immune system. Follow-up studies in normal subjects after cessation of smoking. *The Medical Journal of Australia*, *2*, 425–429.

Jia, H., Uphold, C.R., Zheng, Y., Wu, S., Chen, G.J., Findley, K., & Duncan, P.W. (2007). A further investigation of health-related quality of life over time among men with HIV infection in the HAART era. *Quality of Life Research*, *16*, 961–968.

Kalra, R., Singh, S.P., Savage, S.M., Finch, G.L., & Sopori, M.L. (2000). Effects of cigarette smoke on immune response: Chronic exposure to cigarette smoke impairs antigen-mediated signaling in T cells and depletes IP3- sensitive $Ca2^{+}$ stores. *The Journal of Pharmacology and Experimental Therapeutics*, *293*, 166–171.

Kuniak, M.P., Tsunoda, M., Weiss, S.T., Satoh, T., Guevarra, L., & Tollerud, D.J. (1995). Cigarette smoking and serum cytokine levels. *American Journal of Respiratory and Critical Care Medicine*, *15*, A5–A65.

Lesaffre, E., Rizopoulos, D., & Tsonaka, R. (2007). The logistic transform for bounded outcome scores. *Biostatistics*, *8*, 72–85.

Miller, L.G., Goldstein, G., Murphy, M., & Ginns, L.C. (1982). Reversible alterations in immunoregulatory T cells in smoking. *Chest*, *5*, 527–529.

Niaura, R., Shadel, W.G., Morrow, K., Tashima, K., Flanigan, T., & Abrams, D.B. (2000). Human immunodeficiency virus infection, AIDS, and smoking cessation: The time is now. *Clinical Infectious Diseases*, *31*, 808–812.

Nieman, R.B., Fleming, J., Coker, R.J., William, H.J.R., & Mitchell, D.M. (1993). The effect of cigarette smoking on the development of AIDS in HIV-1-seropositive individuals. *AIDS*, *7*, 705–710.

Okuyemi, K.S., Ahluwalia, J.S., Banks, R., Harris, K.J., Mosier, M.C., Nazir, N., & Powell, J. (2004). Differences in smoking and quitting experiences by levels of smoking among African Americans. *Ethnicity & Disease*, *14*, 127–133.

Okuyemi, K.S., Harris, K.J., Scheibmeir, M., Choi, W.S., Powell, J., & Ahluwalia, J.S. (2002). Light smokers: Issues and recommendations. *Nicotine & Tobacco Research*, *4*, S103–S112.

Ouyang, Y., Virasch, N., Hao, P., Aubrey, M.T., Mukerjee, N., Bierer, B.E., & Freed, B.M. (2000). Suppression of human IL-1β, IL-2, IFN-γ, and TNF-α production by cigarette smoke extracts. *Journal of Allergy and Clinical Immununology*, *106*, 280–287.

Pachl, C., Todd, J.A., Kern, D.G., Sheridan, P.J., Fong, S., Stempien, M., & Urdea, M.S. (1995). Rapid and precise quantification of HIV-1 RNA in plasma using a branched DNA signal amplification assay. *Journal of Acquired Immune Deficiency Syndromes*, *8*, 446–454.

Palacio, H., Hilton, J.F., Canchola, A.J., & Greenspan, D. (1997). Effect of cigarette smoking on HIV-related oral lesions. *Journal of Acquired Immune Deficiency Syndromes*, *14*, 338–342.

Palella, F.J., Delaney, K.M., Moorman, A.C., Loveless, M.O., Fuhrer, J., Satten, G.A., ... Holmberg, S.D. (1998). Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *New England Journal of Medicine*, *338*, 853–860.

Petersen, B., Steimel, L., & Callaghan, J. (1983). Suppression of mitogen-induced lymphocyte transformation in cigarette smokers. *Clinical Immunology and Immunopathology*, *27*, 35–140.

Radloff, L.S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.

Robbins, C.S., Dawe, D.E., Goncharova, S.I., Pouladi, M.A., Drannik, A.G., Swirski, F.K., & Stampfli, M.R. (2004). Cigarette smoke decreases pulmonary dendritic cells and impacts antiviral immune responsiveness. *American Journal of Respiratory Cell and Molecular Biology*, *30*, 202–211.

Robbins, C.S., Franco, F., Mouded, M., Cernadas, M., & Shapiro, S.D. (2008). Cigarette smoke exposure impairs dendritic cell maturation and T cell proliferation in thoracic lymph nodes of mice. *The Journal of Immunology*, *180*, 6623–6628.

Royce, R.A., & Winkelstein, W., Jr. (1990). HIV infection, cigarette smoking and $CD4^+$ T-lymphocyte counts: Preliminary results from the San Francisco Men's Health Study. *AIDS*, *4*, 327–334.

Samet, J.H., Cheng, D.M., Libman, H., Nunes, D.P., Alperen, J.K., & Saitz, R. (2007). Alcohol consumption and HIV disease progression. *Journal of Acquired Immune Deficiency Syndromes*, *46*, 194–199.

Samet, J.H., Horton, N.J., Meli, S., Freedberg, K.A., & Palepu, A. (2004). Alcohol consumption and antiretroviral adherence among HIV-infected persons with alcohol problems. *Alcoholism: Clinical and Experimental Research*, *28*, 572–577.

Samet, J.H., Horton, N.J., Traphagen, E.T., Lyon, S.M., & Freedberg, K.A. (2003). Alcohol consumption and HIV disease progression: Are they related? *Alcoholism: Clinical and Experimental Research*, *27*, 862–867.

Samet, J.H., Libman, H., LaBelle, C., Steger, K., Lewis, R., Craven, D.E., & Freedberg, K.A. (1995). A model clinic for the initial evaluation and establishment of primary care for persons infected with human immunodeficiency virus. *Archives of Internal Medicine*, *155*, 1629–1633.

Samet, J.H., Phillips, S.J., & Horton, N.J. (2004). Detecting alcohol problems in HIV-infected patients. Use of the CAGE questionnaire. *AIDS Research and Human Retroviruses*, *20*, 151–155.

Silverman, N., Potvin, C., Alexander, J.C., Jr. & Chretien, P.B. (1975). In vitro lymphocyte reactivity and T Cell levels in chronic cigarette smokers. *Clinical & Experimental Immunology*,, *22*, 285–292.

Slavinsky, J., III., Myers, T., Swoboda, R.K., Leigh, J.E., Hager, S., & Fidel, P.L. Jr., (2002). Th1/Th2 cytokine profiles in saliva of HIV-positive smokers with oropharyngeal candidiasis. *Oral Microbiology and Immunology*, *17*, 38–43.

Smith, K.L., Horton, N.J., Saitz, R., & Samet, J.H. (2006). The use of the mini-mental state examination in recruitment for substance abuse research studies. *Drug and Alcohol Dependency*, *82*, 231–237.

Sopori, M.H., Gairola, C.C., DeLucia, A.J., Bryant, L.R., & Cherian, S. (1985). Immune responsiveness of monkeys exposed chronically to cigarette smoke. *Clinical Immunology and Immunopathology*, *36*, 338–344.

Sopori, M.L., & Kozak, W. (1998). Immunomodulatory effects of cigarette smoke. *Journal of Neuroimmunology*, *83*, 148–156.

Stephenson, J.M., Griffioen, A., Woronowski, H., Phillips, A.N., Petruckevitch, A., Keenlyside, R., & Doyle, C. (1999). Survival and progression of HIV disease in women attending GUM/HIV clinics in Britain and Ireland. *Sexually Transmitted Infections*, *75*, 247–252.

Sunyer, J., Munoz, A., Peng, Y., Margolick, J., Chmiel, J.S., Oishi, J., ... Samet, J.M. (1996). Longitudinal relation between smoking and white blood cells. *American Journal of Epidemiology*, *144*, 734–741.

Thomas, W., Holt, P., & Keast, D. (1975). Humoral immune response of mice with long-term exposure to cigarette smoke. *Archives of Environmental Health*, *30*, 78–80.

Tollerud, D.J., Clark, J.W., Brown, L.M., Neuland, C.Y., Mann, D.L., Pankiw-Trost, L.K., & Hoover, R.N. (1989a). The effects of cigarette smoking on T cell subsets. A population-based survey of healthy Caucasians. *American Journal of Respiratory and Critical Care Medicine*, *139*, 1446–1451.

Tollerud, D.J., Clark, J.W., Brown, L.M., Neuland, C.Y., Mann, D.L., Pankiw-Trost, L.K., & Hoover, R.N. (1989b). Association of cigarette smoking with decreased numbers of circulating natural killer cells. *The American Review of Respiratory Disease*, *139*, 194–196.

Webb, M.S., Vanable, P.A., Carey, M.P., & Blair, D.C. (2007). Cigarette smoking among HIV+ men and women: Examining health, substance use, and psychosocial correlates across the smoking spectrum. *Journal of Behavioral Medicine*, *30*, 371–383.

Wewers, M.D., Diaz, P.T., Wewers, M.E., Lowe, M.P., Nagaraja, H.N., & Clanton, T.L. (1998). Cigarette smoking in HIV infection induces a suppressive inflammatory environment in the lung. *American Journal of Respiratory and Critical Care Medicine*, *158*, 1543–1549.

Webber, P.M., Schoenbaum, E.E., Gourevitch, M.N., Buono, D., & Klein, R.S. (1999). A prospective study of HIV disease progression in female and male drug users. *AIDS*, *13*, 257–262.

Original Article

Minimal Social Network Effects Evident in Cancer Screening Behavior

Nancy L. Keating, MD, MPH[1,2]; A. James O'Malley, PhD[2]; Joanne M. Murabito, MD, ScM[3,4]; Kirsten P. Smith, PhD[2]; and Nicholas A Christakis, MD, PhD[2,5,6]

BACKGROUND: Social networks may influence screening behaviors. We assessed whether screening for breast, prostate, or colorectal cancer is influenced by the actual screening behaviors of siblings, friends, spouses, and coworkers. **METHODS:** We conducted an observational study using Framingham Heart Study data to assess screening for eligible individuals during the late 1990s. We used logistic regression to determine whether the probability of screening for breast, prostate, or colorectal cancer was influenced by the proportion of siblings, friends, and coworkers who had the same screening, as well as spouse's screening for colorectal cancer, adjusting for other factors that might influence screening rates. **RESULTS:** Among 1660 women aged 41-70 years, 71.7% reported mammography in the previous year; among 1217 men aged 51-70 years, 43.3% reported prostate-specific antigen testing in the previous year; and among 1426 men and women aged 51-80 years, 46.9% reported stool blood testing and/or sigmoidoscopy in the previous year. An increasing proportion of sisters who had mammography in the previous year was associated with mammography screening in the ego (odds ratio [OR], 1.034; 95% confidence interval [CI], 1.000-1.065 for each 10% increase). A spouse with recent screening was associated with more colorectal cancer screening (OR, 1.65; 95% CI, 1.39-1.98 vs unmarried). Otherwise, screening behaviors of siblings, friends, and coworkers were not associated with screening in the ego. **CONCLUSIONS:** Aside from a slight increase in breast cancer screening among women whose sisters were screened and colorectal cancer screening if spouses were screened, the screening behavior of siblings, friends, or coworkers did not influence cancer screening behaviors. ***Cancer* 2011;117:3045-52.** © *2011 American Cancer Society.*

KEYWORDS: cancer screening, mammography, prostate specific antigen, social networks.

Screening for cancer has the potential to save lives by identifying cancers at earlier stages, when they may be more amenable to treatment and cure. Nevertheless, many individuals who may benefit do not undergo routine screening.[1-5] Research suggests that individuals' social support networks, including family and friends, or their perception that screening is normative among their peers, may positively encourage screening.[6-12] Such findings have led to interventions using peers and/or other community members or worksite interventions to increase rates of screening.[13-17]

Social contacts can strongly influence a variety of behaviors, including smoking, weight gain, and drinking.[18-20] They might also influence screening behaviors by several mechanisms. Social contacts might provide information or advice about the purpose of specific tests, the benefits of testing, or the need for evaluation of symptoms. They might also provide encouragement to someone who has avoided screening and may provide emotional support to someone concerned about abnormal screening results. They might share their own experiences with screening, which may be more powerful at influencing behavior than sharing less direct knowledge about the tests. Finally, social contacts might assist an individual in finding a doctor or getting to appointments.

Although other studies have suggested benefits of generic social support in encouraging screening behaviors, we are unaware of studies examining whether actual screening behavior of one's peers influence an individual's likelihood of cancer screening. We studied a large network of individuals to assess whether screening for breast cancer, prostate cancer, or

Corresponding author: Nancy L. Keating, MD, MPH, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115; Fax: (617) 432-0173; keating@hcp.med.harvard.edu

[1]Division of General Internal Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts; [2]Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts; [3]National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, Massachusetts; [4]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts; [5]Department of Medicine, Harvard Medical School, Boston, Massachusetts; [6]Department of Sociology, Harvard University, Cambridge, Massachusetts

We thank Laurie Meneades for expert programming assistance and Garrett Kirk for administrative assistance.

DOI: 10.1002/cncr.25849, **Received:** August 11, 2010; **Revised:** October 15, 2010; **Accepted:** November 16, 2010, **Published online** January 24, 2011 in Wiley Online Library (wileyonlinelibrary.com)

colorectal cancer is influenced by the actual screening behaviors of one's siblings, friends, and coworkers, who may differ in their likelihood of discussing screening and their influence on the ego. We hypothesized that the screening behavior of individuals would be positively influenced by the screening behaviors of their contacts, with the effects being greater for siblings and friends than coworkers. In this study, we were able to measure actual behavior of individuals' social contacts, rather than rely on an index case to inform us about the behavior of others to whom they were connected.

MATERIALS AND METHODS

Data and Subjects

This study used data from the Framingham Heart Study, which, in 1948, enrolled 5209 individuals in the original cohort.[21] Children of the original cohort and their spouses were recruited in 1971 to form the offspring cohort; this cohort included 5124 individuals.[22] In 1994, a minority oversample of 508 individuals was initiated, and in 2002, the third-generation cohort, consisting of 4095 children of the offspring cohort, was initiated.[23] The study protocol was approved by the institutional review boards of Harvard Medical School and Boston University Medical Center.

Network

We focused on the 3807 so-called "egos" in the offspring cohort known to have participated in waves 6 and/or 7 (data collection occurred during 3-year periods centered in 1997 for wave 6 and 1999 for wave 7). The ego is the person whose screening behavior is being analyzed. Any persons to whom these subjects are linked (in any of the Framingham Heart Study cohorts) can serve as social contacts, referred to as "alters." Details of the ascertainment of alters are described in detail elsewhere.[18] Briefly, information was derived from archived, administrative tracking sheets used to identify people close to the study participants to facilitate follow-up. The tracking sheets provided complete information about all first-order relatives (parents, spouses, siblings, and children) and at least 1 "close friend," and these names were linked with the Framingham data to identify ties between egos and alters. Information on address and place of work were used to identify neighbors and coworkers. We classified alters as parents, full sisters, full brothers, friends, and coworkers. We restricted analyses to individuals and their alters who were eligible for screening based on sex- and age-specific screening recommendations (described below). In sensitivity analyses, we repeated all analyses including a small number of half siblings, step siblings, adopted siblings, and foster siblings (5.8% of all siblings). Results were similar and are not presented. We also identified spouses for analyses of colorectal cancer screening.

Screening Behaviors

Data on screening were collected in 2 waves: wave 6 (1995-1998) and wave 7 (1998-2001). Women were asked the year of their last mammogram, men were asked the year of their last blood test for prostate cancer, and men and women were asked the year when stool was last tested for blood and when sigmoidoscopy was last performed.

We assessed screening behavior based on recommendations from national guidelines. For breast cancer screening, we assessed report of mammography within the previous year for women ages 41-70 years[24,25] (results were similar when we restricted to ages 51-70 years, because some guidelines did not recommend mammography for average risk women in their 40s before 1997[26] or later,[27,28] and results were similar when we assessed mammography within the last 2 years). For prostate cancer screening, we assessed prostate-specific antigen (PSA) testing within the previous year among men aged 51-70 years.[25,29] For colorectal cancer screening, we assessed receipt of stool blood testing within the previous year and/or sigmoidoscopy within the past 5 years for individuals aged 51-80 years. Screening guidelines at the time recommended yearly stool occult blood testing or flexible sigmoidoscopy every 3 to 5 years with stool blood testing or colonoscopy;[25,27,30] the survey did not ask specifically about colonoscopy.

Control Variables

We identified factors likely to be associated with cancer risk and/or screening behavior. Specifically, we documented each participant's sex and age and we used data from the prior survey wave (including wave 5 information for subjects whose screening behavior was ascertained in wave 6) to characterize additional participant characteristics, including self-rated health status (excellent/very good, good, fair/poor, unknown), number of years of education, marital status, number of children living (whether in the Framingham Study or not), current employment, number of times per week of intense physical activity, current smoking status, number of alcoholic drinks per week, and presence of diabetes, cardiovascular disease, and/or

pulmonary disease (asthma or chronic obstructive pulmonary disease/emphysema). We used validated data from the Framingham Study to document history of cardiovascular disease and diabetes; presence of pulmonary disease was based on the clinical impression of the Framingham Study clinic examiner. Continuous variables were not categorized.

Analyses

We evaluated testing among eligible participants for each of the waves; hence participants eligible in both waves contributed 2 sets of data to the analyses. We used logistic regression models to assess the proportion of siblings, friends, and coworkers who had the same type of screening test on screening for eligible egos (calculating the effect for each 10% increase in the proportion screened). We also controlled for the number of siblings, friends, and coworkers who were eligible for screening. Models for mammography were limited to female siblings, friends, and coworkers, and those for PSA testing were limited to male siblings, friends, and coworkers. Models also included all control variables described above and survey wave (wave 7 vs 6). The colorectal cancer screening model included men and women, and we included a variable for sex and a variable reflecting marital status and if married whether spouse was screened or not screened. All models used generalized estimating equations, clustering on participants, to account for the possibility that a participant may contribute up to 2 dependent variable observations (1 each wave in the role of ego) or be involved in multiple observations of the predictor variable (in the role of alter).

RESULTS

We identified 1660 women aged 41-70 years, who had 597 sisters, 175 female friends, and 174 female coworkers aged 41-70 years enrolled in the Framingham Heart Study. A total of 1269 women participated in both waves of the survey, so the total number of observations was 2929 women eligible for mammography; of these 71.7% had undergone mammography in the previous year. We identified 1217 men aged 51-70 years, who had 337 brothers, 142 male friends, and 99 male coworkers aged 51-70 years. A total of 804 men participated in both waves of the survey, so the total number of observations was 2021; 43.3% had undergone PSA testing in the previous year. We identified 3045 men and women aged 51-80 years. These individuals had 1426 siblings, 364 friends, and 299 coworkers aged 51-80 years, and 1530 had spouses aged 51-80 years. A total of 2260 individuals participated in both waves, so the total number of observations was 5305; 46.9% had undergone stool blood testing in the previous year and/or flexible sigmoidoscopy in the last 5 years. Characteristics of each cohort are included in Table 1.

Table 2 demonstrates the influence of the screening behaviors of siblings, friends, and coworkers on screening behavior of the ego. For mammography screening, an increasing proportion of sisters who had undergone screening mammography in the previous year was slightly associated with mammography screening in the ego (odds ratio [OR], 1.034; 95% confidence interval [CI], 1.000-1.065 for each 10% increase in the proportion of sisters screened). At an average rate of 71.7% of women being screened, this OR corresponds to a risk ratio[31] of 1.009, suggesting a very small 0.9% increase in screening rates to 72.4% for a 10% increase in the proportion of sisters screened. Women with a greater number of sisters were less likely than women with fewer sisters to undergo mammography (OR, 0.85; 95% CI, 0.72-1.01), although this finding was of borderline statistical significance ($P = .06$). The proportion of female friends and female coworkers who had undergone mammography was not associated with the probability of the ego undergoing mammography screening.

The extent of PSA testing in the previous year among siblings, friends, and coworkers was not associated with PSA testing among egos, nor was the number of brothers, male friends, or male coworkers.

For colorectal cancer screening with stool blood tests and/or sigmoidoscopy, individuals married to a spouse that had been screened were more likely to be screened than those who were unmarried (OR, 1.65; 95% CI, 1.39-1.98), with a risk ratio[31] of 1.296. With 42.6% of unmarried individuals screened, this corresponds to a 12.6% absolute increased risk of screening, to 55.2% for married individuals whose spouses were screened. Individuals married to a spouse who had not been screened or for whom screening status was unknown did not differ from unmarried individuals in screening. The non-overlapping CIs for married patients whose spouses were or were not screened suggest that screening status of the spouse is a more important factor than marital status itself. The proportion of siblings, friends, or coworkers who had been screened was not associated with screening. Individuals with more friends were less likely to undergo colorectal cancer screening (OR, 0.73; 95% CI, 0.58-0.91 for each additional friend in the cohort).

Table 1. Characteristics of the Study Cohorts

	Breast Cancer Screening Cohort[a]	Prostate Cancer Screening Cohort[b]	Colorectal Cancer Screening Cohort[c]
Mean age, y (SD)	57.2 (7.3)	59.8 (5.5)	62.3 (7.6)
Mean no. of years of education (SD)	13.9 (2.2)	14.6 (2.8)	14.0 (2.5)
Married, %	72.1	85.1	75.9
Spouses screened among married, %			
Unmarried	—	—	24.3
Married, spouses screened	—	—	24.1
Married, spouses not screened	—	—	27.5
Married, unknown if spouses screened	—	—	24.1
Sex	—	—	
Women	—	—	54.0
Men	—	—	46.0
Mean no. of children (SD)	2.6 (1.6)	2.8 (1.6)	2.9 (1.6)
Mean no. of times per week intense physical activity (SD)	2.2 (2.1)	2.8 (2.7)	2.4 (2.4)
Currently working, %	61.7	64.7	51.8
Mean no. of drinks per week (SD)	3.6 (5.4)	7.6 (10.0)	5.2 (7.9)
Current smoker, %	16.5	18.3	15.4
Self-reported health, %			
Excellent/very good	43.4	43.9	39.9
Good	49.9	47.9	51.2
Fair/poor	6.0	7.3	7.7
Unknown	0.7	0.9	1.1
No. of comorbidities, %[d]			
0	83.9	76.6	78.1
1	15.1	20.2	19.3
2 or 3	1.0	3.2	2.6
Survey wave, %			
Wave 6	51.2	50.1	47.7
Wave 7	48.8	49.9	52.3
Mean no. of eligible siblings in cohort[e] (SD), range	0.4 (.7), 0-4	0.3 (0.6), 0-3	0.7 (1.0), 0-5
Mean no. of friends in cohort (SD)	0.1 (0.3), 0-2	0.1 (0.3), 0-1	0.1 (0.3), 0-2
Mean no. of coworkers in cohort (SD)	0.2 (0.9), 0-10	0.1 (0.5), 0-7	0.2 (0.9), 0-10
% reporting screening in previous year	71.7	43.3	46.9

[a] Based on 2929 observations for 1660 women.
[b] Based on 2021 observations for 1217 men.
[c] Based on 5305 observations for 3045 men and women.
[d] Considering heart disease, diabetes, and chronic obstructive pulmonary disease.
[e] Siblings included sisters for breast cancer screening cohort, brothers for prostate cancer screening cohort, and sisters and brothers for colorectal cancer screening cohort who met screening criteria.

In each model, we adjusted for several control variables, some of which had significant effects. Other characteristics of the egos associated with mammography screening included older age (OR, 1.02; 95% CI, 1.01-1.04 for each year of age), being married (OR, 1.63; 95% CI, 1.33-2.00), and being physically active 4 or more times per week (OR, 1.05; 95% CI, 1.00-1.09). Smokers were much less likely to undergo mammography than nonsmokers (OR, 0.56; 95% CI, 0.44-0.71), as were women with more comorbid illnesses (OR, 0.80; 95% CI, 0.65-0.99). Participants surveyed in the wave centered in 1999 were more likely to have mammograms than those surveyed in the wave centered in 1997 (OR, 1.40; 95% CI, 1.21-1.62).

For PSA testing, older men were more likely to be screened than younger men (OR, 1.05; 95% CI, 1.03-1.08). Married men had more PSA tests than unmarried men (OR, 1.73; 95% CI, 1.29-2.31), and men with more years of education were more likely to have PSA testing (OR, 1.08; 95% CI, 1.04-1.12). Smokers were less likely than nonsmokers to have PSA testing (OR, 0.71; 95% CI, 0.54-0.94), as were men with more comorbid illnesses (OR, 0.76; 95% CI, 0.62-0.94). Men surveyed in wave 7 were more likely to report recent PSA testing than those

Table 2. Factors Associated With Recommended Screening in Egos

Alters Use of Screening	Mammogram in Previous Year	Prostate-Specific Antigen Test in Previous Year	Stool Card in Previous Year and/or Flexible Sigmoidoscopy in Previous 5 Years
Siblings[a]			
10% increase in proportion of siblings with test	1.034 (1.000-1.065)[b]	1.005 (0.969-1.042)	1.009 (0.992-1.027)
No. of siblings	0.85 (0.72-1.01)[c]	0.87 (0.72-1.07)	0.95 (0.89-1.02)
Friends			
10% increase in proportion of friends with test	1.012 (0.958-1.070)	1.021 (0.970-1.074)	1.017 (0.986-1.049)
No. of friends	0.96 (0.57-1.64)	1.08 (0.73-1.59)	0.73 (0.58-0.91)[b]
Coworkers			
10% increase in proportion of coworkers with test	0.987 (0.941-1.037)	0.947 (0.803-1.021)	1.023 (0.984-1.064)
No. of coworkers	0.96 (0.86-1.07)	1.11 (0.89-1.38)	0.97 (0.89-1.05)
Spouse screened among married, %			
Unmarried	—	—	1.0
Married, spouse screened	—	—	1.66 (1.39-1.98)[b]
Married, spouse not screened	—	—	1.06 (0.89-1.27)
Married, unknown if spouse screened	—	—	1.07 (0.89-1.27)
Age	1.02 (1.01-1.04)[b]	1.05 (1.03-1.08)[b]	1.02 (1.01-1.03)[b]
Married	1.63 (1.33-2.00)[b]	1.73 (1.29-2.31)[b]	—
Sex			
Men	—	—	1.0
Women	—	—	1.00 (0.87-1.15)
No. of years of education	0.99 (0.94-1.03)	1.08 (1.04-1.12)[b]	1.06 (1.03-1.09)[b]
No. of children	0.99 (0.92-1.05)	0.95 (0.89-1.02)	1.00 (0.96-1.04)
Currently working	0.90 (0.73-1.10)	0.90 (0.71-1.14)	0.85 (0.74-0.98)[b]
Physical activity	1.05 (1.00-1.09)[b]	1.00 (0.96-1.04)	1.00 (0.98-1.02)
No. of drinks per week	1.00 (0.98-1.02)	1.00 (0.99-1.01)	1.00 (1.00-1.01)
Smoker	0.56 (0.44-0.71)[b]	0.71 (0.54-0.94)[b]	0.62 (0.51-0.74)[b]
Self-reported health status			
Very good/excellent	1.0	1.0	1.0
Good	1.01 (0.84-1.21)	0.95 (0.78-1.15)	0.97 (0.86-1.09)
Fair/poor	1.09 (0.74-1.61)	0.85 (0.56-1.27)	1.04 (0.82-1.32)
Unknown	0.37 (0.15-0.95)[b]	0.64 (0.21-1.94)	0.37 (0.20-0.69)[b]
No. of comorbid illnesses[d]	0.80 (0.65-0.99)[b]	0.76 (0.62-0.94)	0.83 (0.73-0.94)
Survey wave			
Wave 6	1.0	1.0	1.0
Wave 7	1.40 (1.21-1.62)	1.59 (1.35-1.87)	1.52 (1.38-1.67)

All data are presented as odds ratio (95% confidence interval). Values were calculated using generalized estimating equations to account for clustering within participants, because some participants had data from both waves 6 and 7 and/or functioned as alters multiple times. Data were adjusted for participant age, level of education, marital status, number of children, employment status, level of physical activity, smoking status, self-reported health status, coronary heart disease, diabetes, chronic obstructive pulmonary disease, and survey wave. Only female siblings, friends, and coworkers were included in mammography analyses; only male siblings were included in prostate-specific antigen analysis.

[a] Siblings included sisters for mammography analysis, brothers for prostate-specific antigen analysis, and sisters and brothers for colorectal screening analysis who met screening criteria.

[b] $P < .05$.

[c] $P < .10$.

[d] Considering heart disease, diabetes, and chronic obstructive pulmonary disease.

surveyed in wave 6 (OR, 1.59; 95% CI, 1.35-1.87), reflecting the general increase in the use of screening over this time frame.

For stool blood testing or flexible sigmoidoscopy, older participants were more likely to report screening (OR, 1.02; 95% CI, 1.01-1.03). Participants who worked were less likely than those who did not to be screened (OR, 0.85; 95% CI, 0.74-0.98), as were those who smoked (OR, 0.62; 95% CI, 0.51-0.74) and those with more comorbidities (OR, 0.83; 95% CI, 0.73-0.94). Participants surveyed in the wave centered in 1999 were more likely to undergo colon cancer screening than those surveyed in the wave centered in 1997 (OR, 1.52; 95% CI, 1.38-1.67).

DISCUSSION

We examined whether screening behavior of siblings, friends, coworkers, and spouses influenced analogous screening behaviors of individuals. We found that mammography screening increases slightly with an increasing proportion of sisters who have had a mammogram (although women with more sisters were less likely to get mammograms, a finding of borderline statistical significance). PSA testing did not vary by the proportion of brothers, friends, or coworkers who had the test. Colorectal cancer screening was strongly associated with screening among one's spouse, but not with the proportion of friends who were screened.

Several studies suggest that support of others increases an individual's likelihood of participating in cancer screening. For example, women with higher scores on the social network index[6,7] or who report social support from physicians, family, and friends[8] are more likely to undergo mammography and Pap smears. In addition, a study of employed women found that women who perceived that screening is normative among their peers were more likely to undergo regular mammography, although the extent of social support and the size of one's social network was not associated with screening behavior.[9] On the other hand, 2 studies have observed that women reporting explicit encouragement to undergo mammography by social network members were less likely to be screened,[9,32] suggesting that the women who avoid mammography may be more likely to be offered encouragement from others. Perceived risk of cancer is also associated with mammography screening, and this perceived risk is often due to a family history of cancer.[33] In a previous study using data from the Framingham Study, reporting a family history of breast cancer was strongly associated with reporting a mammogram in the last 2 years.[34]

In the current study, we were able to broaden the scope of social contacts examined (to include friends and coworkers), broaden the nature of cancers considered, and, most importantly, trace out direct ties between people and directly query alters about screening behavior rather than merely surveying egos about alters.

Past research suggests that friends can influence mammography behavior with direct efforts. One study randomized individuals to call or not call friends to encourage them to get a mammogram. Friends who received a call had a 15% increase in mammography compared with those who did not receive a call. This effect remained after controlling for demographic characteristics, was effective for black and white women of all ages, and was most pronounced among women with lower household incomes.[35] In addition, women reporting close friends with whom they could discuss their health were more likely to have ever had a Pap smear.[11] Programs have thus been developed that successfully use social support to improve screening for cervical cancer and breast cancer.[36]

Fewer data are available about the impact of interventions on social contacts on prostate cancer screening or colorectal cancer screening. The value of PSA testing for prostate cancer remains controversial,[37] yet research suggests that patients deciding about PSA testing value anecdotes about the decisions of friends, family, or media celebrities.[38] Thus, we had expected that the prostate screening behaviors of alters would influence those of the egos in our study. Consistent with other research,[39] we found that married men were more likely than unmarried men to undergo PSA screening. Men may be encouraged by their wives to obtain more routine and preventive care, or may be persuaded to undergo PSA testing specifically.

Colorectal cancer screening can be inconvenient and invasive, and, for colonoscopy, requires time off from work and someone to accompany the individual to the procedure. These factors may lead to negative attitudes about screening.[40] Nevertheless, support from friends and family has been associated with screening, as have positive attitudes about the screening and beliefs that it is safe.[41] We found a strong association of colorectal cancer screening among spouses of individuals who have been screened, but no associations based on the proportion of siblings, friends, or coworkers who were screened.

Overall, this work again reinforces the distinction between social support and social network effects.[42] The existence of social ties, and the willingness of others to help with health care can affect screening, as suggested by prior work. However, this is a different effect than that of the specific influence whereby an alter's actual behavior influences a similar behavior in an ego. By analogy, it is the difference between the impact on a person's happiness of having many friends versus the impact on a person's happiness of having friends who are themselves happy.[43] We found that the screening behaviors of one's contacts, at least among those contacts included in the study, had little relevance to screening behaviors. Screening behaviors may be less "contagious" because they are not easily observed (unlike smoking, alcohol, obesity, and happiness)[18-20,43] and may not be comfortable topics to discuss. New evidence suggests that ties among friends are influenced by observable characteristics such as obesity and

smoking, but not by less easily observed traits, such as blood pressure and depression score.[44]

Our findings should be interpreted in light of some limitations. First, information on screening was only collected in 2 waves of the Framingham Study, both during the late 1990s. Consequently, we cannot be certain that the findings are relevant to current screening behaviors; screening rates for colorectal cancer have increased since this time,[45] although our study period corresponded with the peaking of mammography rates, which declined in the early 2000s.[46] Second, our study focused on a single community that was lacking in racial and ethnic diversity, so the generalizability of our findings to other populations requires further study. Rates of prostate cancer screening in our cohort were lower than those of colorectal cancer screening, which has not been observed nationally.[47] Third, we could only assess screening behaviors among alters who were included in the Framingham cohorts and of ages that would make them eligible for screening themselves, and our cohort of egos had relatively few alters in the study, limiting our ability to observe effects. Moreover, if an individual had many friends but few were in the Framingham cohorts, then our study would likely underestimate the effects of the other friends' behaviors. Fourth, the survey question about stool blood testing did not distinguish in-office or at-home testing. Finally, self-report of screening may overestimate use.[48]

In conclusion, mammography receipt among sisters and colorectal screening among spouses had some influence on personal screening behaviors, but otherwise screening behaviors of siblings, friends, and coworkers were not associated with increased rates of cancer-specific screening. These observations suggest that while many health behaviors may spread across social ties, not all health behaviors necessarily do. Some behaviors may be intrinsically more "contagious," just as some fashions are easier to adopt and some germs are more contagious than others.

CONFLICT OF INTEREST DISCLOSURES

This study was funded by the National Institute on Aging (grant nos. R-01 AG-24448 and P01 AG-031093) and the National Heart, Lung, and Blood Institute Framingham Heart Study (contract no. N01-HC-25195). N.A.C. has an equity stake in a company, MedNetworks, that is licensed by Harvard to apply some of the ideas embodied in work on social networks.

REFERENCES

1. Rimer BK, Conaway MR, Lyna PR, et al. Cancer screening practices among women in a community health center population. *Am J Prev Med*. 1996;12:351-357.
2. Cooper GS, Doug Kou T. Underuse of colorectal cancer screening in a cohort of Medicare beneficiaries. *Cancer*. 2008;112:293-299.
3. Ko CW, Kreuter W, Baldwin LM. Persistent demographic differences in colorectal cancer screening utilization despite Medicare reimbursement. *BMC Gastroenterol*. 2005;5:10.
4. McBean AM, Yu X. The underuse of screening services among elderly women with diabetes. *Diabetes Care*. 2007; 30:1466-1472.
5. Demark-Wahnefried W, Strigo T, Catoe K, et al. Knowledge, beliefs, and prior screening behavior among blacks and whites reporting for prostate cancer screening. *Urology*. 1995;46:346-351.
6. Kang SH, Bloom JR, Romano PS. Cancer screening among African-American women: their use of tests and social support. *Am J Public Health*. 1994;84:101-103.
7. Suarez L, Lloyd L, Weiss N, Rainbolt T, Pulley L. Effect of social networks on cancer-screening behavior of older Mexican-American women. *J Natl Cancer Inst*. 1994;86:775-779.
8. Taylor VM, Thompson B, Montano DE, Mahloch J, Johnson K, Li S. Mammography use among women attending an inner-city clinic. *J Cancer Educ*. 1998;13:96-101.
9. Allen JD, Sorensen G, Stoddard AM, Peterson KE, Colditz G. The relationship between social network characteristics and breast cancer screening practices among employed women. *Ann Behav Med*. 1999;21:193-200.
10. Suarez L, Ramirez AG, Villarreal R, Marti J, McAlister A, Talavera GA, et al. Social networks and cancer screening in four U.S. Hispanic groups. *Am J Prev Med*. 2000;19:47-52.
11. Seow A, Huang J, Straughan PT. Effects of social support, regular physician and health-related attitudes on cervical cancer screening in an Asian population. *Cancer Causes Control*. 2000;11:223-230.
12. Allen JD, Stoddard AM, Sorensen G. Do social network characteristics predict mammography screening practices? *Health Educ Behav*. 2008;35:763-776.
13. Stockdale SE, Keeler E, Duan N, Derose KP, Fox SA. Costs and cost-effectiveness of a church-based intervention to promote mammography screening. *Health Serv Res*. 2000;35:1037-1057.
14. Allen JD, Stoddard AM, Mays J, Sorensen G. Promoting breast and cervical cancer screening at the workplace: results from the Woman to Woman Study. *Am J Public Health*. 2001;91:584-590.
15. Derose KP, Duan N, Fox SA. Women's receptivity to church-based mobile mammography. *J Health Care Poor Underserved*. 2002;13:199-213.
16. Katz ML, Tatum C, Dickinson SL, et al. Improving colorectal cancer screening by using community volunteers: results of the Carolinas cancer education and screening (CARES) project. *Cancer*. 2007;110:1602-1610.
17. Paskett E, Tatum C, Rushing J, et al. Randomized trial of an intervention to improve mammography utilization among a triracial rural population of women. *J Natl Cancer Inst*. 2006;98:1226-1237.
18. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med*. 2007;357:370-379.
19. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *N Engl J Med*. 2008; 358:2249-2258.
20. Rosenquist JN, Murabito J, Fowler JH, Christakis NA. The spread of alcohol consumption behavior in a large social network. *Ann Intern Med*. 2010;152:426-433.

21. Dawber T. The Framingham Study: the epidemiology of atherosclerotic disease. Cambridge, MA: Harvard University Press; 1980.
22. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med*. 1975;4:518-525.
23. Splansky GL, Corey D, Yang Q, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol*. 2007;165:1328-1335.
24. ACOG Committee Opinion. Routine cancer screening. Number 185, September 1997 (replaces no. 128, October 1993). Committee on Gynecologic Practice. American College of Obstetricians and Gynecologists. *Int J Gynaecol Obstet*. 1997;59:157-161.
25. Smith RA, Mettlin CJ, Davis KJ, Eyre H. American Cancer Society guidelines for the early detection of cancer. *CA Cancer J Clin*. 2000;50:34-49.
26. National Cancer Institute. National Cancer Advisory Board issues mammography screening recommendations [press release]. Vol. 2009. Bethesda, MD: National Institutes of Health; 1997.
27. U.S. Preventive Services Task Force. Guide to clinical preventive services: report of the U.S. Preventive Services Task Force. 2nd ed. Baltimore, MD: Williams & Wilkins; 1996.
28. U.S. Preventive Services Task Force. Screening for breast cancer: recommendations and rationale. *Ann Intern Med*. 2002;137:344-346.
29. Middleton RG, Thompson IM, Austenfeld MS, et al. Prostate Cancer Clinical Guidelines Panel Summary report on the management of clinically localized prostate cancer. The American Urological Association. *J Urol*. 1995;154:2144-2148.
30. Levin B, Bond JH. Colorectal cancer screening: recommendations of the U.S. Preventive Services Task Force. American Gastroenterological Association. *Gastroenterology*. 1996; 111:1381-1384.
31. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*. 1998;280:1690-1691.
32. Lerman C, Rimer B, Trock B, Balshem A, Engstrom PF. Factors associated with repeat adherence to breast cancer screening. *Prev Med*. 1990;19:279-290.
33. Calvocoressi L, Kasl SV, Lee CH, Stolar M, Claus EB, Jones BA. A prospective study of perceived susceptibility to breast cancer and nonadherence to mammography screening guidelines in African American and White women ages 40 to 79 years. *Cancer Epidemiol Biomarkers Prev*. 2004;13: 2096-2105.
34. Murabito JM, Evans JC, Larson MG, et al. Family breast cancer history and mammography: Framingham Offspring Study. *Am J Epidemiol*. 2001;154:916-923.
35. Calle EE, Miracle-McMahill HL, Moss RE, Heath CW Jr. Personal contact from friends to increase mammography usage. *Am J Prev Med*. 1994;10:361-366.
36. Gotay CC, Wilson ME. Social support and cancer screening in African American, Hispanic, and Native American women. *Cancer Pract*. 1998;6:31-37.
37. Barry MJ. Screening for prostate cancer--the controversy that refuses to die. *N Engl J Med*. 2009;360:1351-1354.
38. Farrell MH, Murphy MA, Schneider CE. How underlying patient beliefs can affect physician-patient communication about prostate-specific antigen testing. *Eff Clin Pract*. 2002;5:120-129.
39. Ross LE, Taylor YJ, Richardson LC, Howard DL. Patterns in prostate-specific antigen test use and digital rectal examinations in the Behavioral Risk Factor Surveillance System, 2002-2006. *J Natl Med Assoc*. 2009;101:316-324.
40. Taskila T, Wilson S, Damery S, et al. Factors affecting attitudes toward colorectal cancer screening in the primary care population. *Br J Cancer*. 2009;101:250-255.
41. Holden D, Harris R, Porterfield D, et al. Enhancing the use and quality of colorectal cancer screening. RTI International-University of North Carolina Evidence-based Practice Center, Contract No. 290-2007-10056-I. Rockville, MD: Agency for Healthcare Research and Quality; 2010.
42. Smith KP, Christakis NA. Social networks and health. *Annu Rev Sociol*. 2008;34:405-429.
43. Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *BMJ*. 2008;337:a2338.
44. O'Malley AJ, Christakis NA. Longitudinal analysis of large social networks: estimating the effect of health traits on changes in friendship ties. *Stat Med*. In press.
45. Centers for Disease Control and Prevention. Increased use of colorectal cancer tests--United States, 2002 and 2004. MMWR Morb Mortal Wkly Rep. 2006;55:308-311.
46. Centers for Disease Control and Prevention. Use of mammograms among women aged $\geq$40, United States, 2000-2005. MMWR Morb Mortal Wkly Rep. 2007;56:49-51.
47. Sirovich BE, Schwartz LM, Woloshin S. Screening men for prostate and colorectal cancer in the United States: does practice reflect the evidence? *JAMA*. 2003;289:1414-1420.
48. Rauscher GH, Johnson TP, Cho YI, Walk JA. Accuracy of self-reported cancer-screening histories: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*. 2008;17:748-57.

Drug and Alcohol Dependence 115 (2011) 80–86

Contents lists available at ScienceDirect

Drug and Alcohol Dependence

journal homepage: www.elsevier.com/locate/drugalcdep

Initiation and engagement in chronic disease management care for substance dependence

Theresa W. Kim[a,*], Richard Saitz[a,b], Debbie M. Cheng[a,b], Michael R. Winter[b], Julie Witas[a], Jeffrey H. Samet[a,b]

[a] Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center, 801 Massachusetts Avenue, Boston, MA 02118, USA
[b] Boston University School of Public Health, 801 Massachusetts Avenue, Boston, MA 02118, USA

ARTICLE INFO

Article history:
Received 12 November 2009
Received in revised form 22 October 2010
Accepted 26 October 2010
Available online 18 December 2010

Keywords:
Performance measures
Chronic disease management
Substance abuse treatment

ABSTRACT

Background: Substance dependence treatment is often episodic and not well coordinated with healthcare for common comorbidities. Chronic disease/care management (CDM), longitudinal, patient-centered care delivered by multidisciplinary health professionals, may be well suited to treat substance dependence (SD).
Objective: To examine initiation and engagement with CDM care for SD located in a primary medical setting.
Methods: We prospectively studied substance dependent participants enrolled in a trial of CDM addiction care. Primary study outcomes, based upon Washington Circle performance measures, were 14-day initiation of CDM care and 30-day engagement with CDM care. Factors associated with these outcomes were determined using multivariable logistic regression models. We also estimated the proportion of participants who eventually attended at least two visits and four visits by the end of the study (Kaplan–Meier method).
Results: Of 282 participants, approximately half of the cohort (45%, 95% Confidence Interval [CI] 39–51%) met criteria for 14-day initiation and 23% (95% CI 18–28%) for 30-day engagement with CDM care. Most participants attended two or more (81%, 95% CI 76–85%) and four or more CDM visits (62%, 95% CI 56–68%). Major depressive episode (AOR 2.60, 95% CI 1.39, 4.87) was associated with higher odds of 14-day initiation; younger age, female sex, and higher alcohol addiction severity were associated with lower odds of 30-day engagement with CDM care.
Conclusion: : People with SD appear to be willing to initiate and engage with CDM care in a primary medical care setting. CDM care has the potential to improve the quality of care for people with addictions.

1. Introduction

Although substance dependence is often an illness with chronic physiological changes and a relapsing course, most addiction treatment is not structured to manage it as a chronic disease (McLellan et al., 2000; Institute of Medicine, 2006). Traditionally, patients are encouraged to enter addiction treatment for a specified period with the unrealistic expectation of "curing" their substance use disorder (SUD). Most people with SUD do not seek specialty treatment (Cunningham and Blomqvist, 2006; Hasin et al., 2007) and for the few who do, treatment is often episodic and prematurely truncated due to dissatisfaction with care, motivational issues, or program challenges (Substance Abuse and Mental Health Services Administration, Office of Applied Studies, 2008). Medical and psychiatric comorbidities are often neglected, despite their potential to interfere with addiction treatment and contribute to relapse. Patients often need help navigating complex systems of care across addiction, medical, and mental health arenas.

To address these shortcomings, the Institute of Medicine (IOM) and others have called attention to the chronic disease management/chronic care model (CDM) (Wagner, 2000) to improve the health care of individuals with chronic illnesses, including SUD (Institute of Medicine, 2006). Though not unequivocally supported, CDM care has shown promise for other chronic conditions including congestive heart failure, chronic pulmonary disease, and depressive disorders (Roy-Byrne et al., 2001; Simon et al., 2001; Rea et al., 2004; Whellan et al., 2005). Though not as yet reported in randomized trials, CDM care for substance dependence (SD) offers the potential to follow individuals longitudinally, monitor disease progression, and enhance treatment adherence (Saitz et al.,

* Corresponding author. Tel.: +1 617 414 6932; fax: +1 617 414 4676.
E-mail address: theresa.kim@bmc.org (T.W. Kim).

0376-8716/$ – see front matter © 2010 Published by Elsevier Ireland Ltd.
doi:10.1016/j.drugalcdep.2010.10.013

2008; McKay, 2009). Similar to treatment of other chronic illnesses, adjustments in treatment intensity and modality can be made based upon a patient's functioning, motivation, and clinical course. Multidisciplinary teams with addiction-specific skills can provide direct care, coordinate referrals, communicate with other clinical caregivers, and proactively arrange use of community resources.

CDM care for SD builds upon growing evidence supporting the effectiveness of continuing care interventions to bridge transitions from more intensive treatment (e.g., residential or intensive outpatient treatment) to less intensive treatment (e.g., group counseling). Post-treatment monitoring is effective for facilitating early readmission to treatment for relapses (Scott and Dennis, 2009). Since considerable effort is often required to maintain ongoing attendance (McKay, 2009), alternative modes of treatment delivery such as telephone-based monitoring and brief counseling have been tested and found to be effective for decreasing continuing care dropout and substance use (McKay, 2005).

Some have suggested that primary care may be an optimal setting for providing CDM care for SD (McKay, 2009). Since primary care is meant to deliver longitudinal care, linking CDM care to primary care may facilitate ongoing utilization of CDM care for periodic assessments. Linking CDM care to primary care offers the potential to identify and treat primary care patients with SD who would otherwise not seek substance abuse treatment (Saitz et al., 2008) and to facilitate the coordinating role of the primary care team with respect to the medical, addiction, and psychiatric systems of care.

Although primary care services are theoretically available to patients with SD, these services are often not received for various reasons, such as missed appointments and lack of follow-through for evaluation of medical problems. CDM located in primary care may facilitate the evaluation of comorbid medical problems by encouraging ongoing attendance and engagement with medical care to facilitate actual receipt of services.

Other potential elements of CDM care that are often not part of "real-world" primary care include a focus on SD as a chronic illness, accessible specialty addiction expertise, and a delivery system designed to facilitate coordination of addiction, medical and psychiatric care (Saitz et al., 2008). CDM care is distinct from primary care and substance abuse treatment in that CDM focuses on increasing utilization of care that is potentially available but often difficult to access. It does so by delivering some of this care directly and by actively facilitating and monitoring access to services outside of the CDM clinic.

Although CDM care for SD is potentially effective, whether this type of care is acceptable to patients with SD is unknown. McLellan specified that "a continuing care approach emphasizes making treatments attractive to patients" (McLellan, 2002). Assessing whether patients find CDM care acceptable is important because efficacious addiction treatments are often underutilized, in part, due to lack of patient acceptance (Tucker et al., 2009). Whether patients will initiate and engage with CDM care is an essential component to assessing its potential effectiveness.

The primary objective of this study was to examine the proportion of study participants that initiated and engaged with CDM addiction care when this modality was made accessible. The secondary objective was to assess characteristics associated with initiation and engagement with CDM addiction care. Initiation and engagement were examined using an adaptation of the Washington Circle (WC) performance measures. These performance measures, adopted by the National Committee for Quality Assurance (NCQA) for inclusion in its Health Plan Employer Data and Information Set (HEDIS) (National Quality Forum, 2007) are associated with beneficial outcomes including lower likelihood of arrests/incarcerations (Garnick et al., 2007) and improvements in alcohol addiction severity (Harris et al., 2010).

2. Materials and methods

2.1. Study design and sample

This is a prospective cohort study of patients with alcohol and/or drug dependence enrolled in the Addiction Health Evaluation and Disease management (AHEAD) study, a randomized controlled trial designed to test the effectiveness of CDM for SD located in primary care. This study's analytic sample included only participants randomized to have access to CDM care. Control participants in the AHEAD study were not included in this study's analysis because they did not have access to CDM addiction care and thus, by design, could neither initiate nor engage with CDM care.

Recruitment for the parent study (the AHEAD randomized trial) occurred at an inpatient detoxification unit, primary care clinics and the emergency department at Boston Medical Center, and from the community by advertising on buses and in newspapers.

Eligible participants were adults with alcohol or drug dependence (Composite International Diagnostic Interview Short Form (CIDI-SF)) (Kessler et al., 1998) and current (past-month) drug (heroin or cocaine) or heavy alcohol use (≥5 drinks per day or >14 drinks per week for men; ≥4 for drinks per day or >7 drinks per week for women) who were willing to establish or continue primary medical care at Boston Medical Center (BMC) and attend an outpatient visit in primary care. Patients who were pregnant, had plans to leave the area or a Mini-mental State Examination score <21 (Smith et al., 2006) were excluded. This study's analysis included all parent study eligibility criteria along with one additional criterion: access to CDM addiction care, defined as randomization to have access to the CDM intervention.

Eligible patients were invited to enroll in a study that may include attending an outpatient visit (i.e., the AHEAD clinic) in a primary medical care clinic. Enrollment of study participants was *not* based upon an interest in utilizing CDM care or any other addiction treatment.

After completing the baseline research interview, participants were accompanied to their first CDM visit in the AHEAD clinic. Participants were compensated for study participation after research assessments and the first (intake) AHEAD clinic visit. Thereafter, AHEAD clinic visits were neither compensated nor required for continued participation in the study. Participants were assessed periodically for research purposes but were neither discouraged nor encouraged to attend the AHEAD clinic by research staff. The Institutional Review Board of Boston University Medical Center approved this study. Additional privacy assurances were secured by the issuance of a Certificate of Confidentiality by the Department of Health and Human Services.

2.2. Description of the chronic disease management (CDM) clinic

The main goals of CDM care were to engage patients in longitudinal addiction treatment tailored to patients' needs (including attention to social, medical, and mental health), to re-engage patients in addiction treatments after relapse and/or loss to clinical follow-up, and to improve addiction-related health outcomes. CDM services included clinical case management, motivational enhancement counseling, addiction pharmacotherapy (i.e., buprenorphine, naltrexone, acamprosate and referral for methadone), psychopharmacology, and referrals for addiction, medical, and psychiatric treatment. Although the AHEAD clinic was located in a large primary medical clinic of an urban "safety-net" hospital, the AHEAD staff did not provide primary care. Instead, the clinic encouraged initiation of primary care and adherence to the evaluation and treatment plan for medical problems. Participants could access medical, psychiatric, and substance abuse services provided by the hospital without referral by the AHEAD clinic.

The AHEAD clinic team was comprised of a nurse care manager, social worker, and physicians with addiction expertise (an internist and a psychiatrist). At the first (intake) visit, the team assessed subjects' addiction, medical, and psychosocial needs and negotiated with the patient to prioritize and address short-term needs. While there were overall clinical guidelines for what each patient should be offered in the AHEAD clinic, participants received different interventions based upon need, availability, and patient preference.

After the intake visit, the nurse care manager tried to maintain periodic contact with patients to provide relapse prevention counseling, address social service concrete needs, and facilitate referrals for care. The clinic allowed patients to attend without appointments, regardless of ongoing substance use. Efforts to encourage follow-up included multiple rescheduling attempts (phone, letter) for patients who missed their appointments.

2.3. Measures

2.3.1. Outcomes. Each of this study's main outcomes, initiation and engagement with CDM care, were derived from the Washington Circle (WC) performance measures for outpatient addiction treatment (Garnick et al., 2002). Because the study's objective was to examine initiation and engagement with CDM care, we adapted these measures to only include CDM visits rather than any outpatient treatment service. Initiation of CDM care was defined as two or more AHEAD visits within 14 days of study entry ("14-day initiation") and engagement with CDM care as two or more AHEAD visits within 30 days of initiation ("30-day engagement"). Even though

a significant portion of the study sample was recruited from detox, we did not use the Washington Circle continuity of care measure after detox because we were specifically interested in initiation and engagement with CDM care rather than the effect of CDM care on continuity of care after a detox admission.

Because participants were not specifically seeking CDM care, we also examined the proportion of the study sample who eventually attended two or more AHEAD visits ("linkage with CDM care") and four or more visits ("continuation of CDM care") over the course of the study. AHEAD clinic attendance was prospectively assessed in a standard fashion using templates specifically created for the clinic in an electronic medical record.

2.3.2. Independent variables. Using Gelberg's vulnerable populations modification of Andersen's behavioral model as a guide for the choice of independent variables, we categorized independent variables, all assessed at study entry, as Need, Enabling, or Predisposing factors (Gelberg et al., 2000). Relevant indicators of need for CDM care were conceptualized into four categories: addiction, social, psychiatric, and medical needs. Addiction-related needs were dependence type (alcohol, drug or both) (CIDI-SF) (Kessler et al., 1998); addiction severity (Addiction Severity Index) (McLellan et al., 1992); history of overdose, and patient-assessed treatment need, which although was not directly measured, was partly reflected by a readiness to change scale.

Social needs included homelessness (any night in a shelter or on the street in the past 3 months) (Kertesz et al., 2006) and current legal problems. Psychiatric needs, assessed with the Mini International Neuropsychiatric Interview (MINI) (Sheehan et al., 1998), were major depressive episode and post-traumatic stress disorder. Medical needs were reflected by two self-report comorbidity questionnaires, one validated by Katz et al. (1996) and another assessing substance-related medical conditions (De Alba et al., 2004).

Enabling variables were: unfavorable social network (most or all of the people that the participant spends time with are either heavy/problem drinkers or drug users); health insurance; and health services utilization of (1) residential addiction treatment (excluding care for detoxification), (2) mutual-help (such as Alcoholics, Cocaine, or Narcotic Anonymous), (3) psychiatric care (counseling or therapy for emotional/psychological problems including full-day treatment, partial hospital program, or treatment by a psychiatrist); and (4) medical hospitalizations.

Predisposing variables included age, sex, and race/ethnicity. To account for potential changes in clinic practices at different points in the study, we included the time of participant enrollment with respect to the first day of cohort recruitment ("study enrollment month").

2.4. Statistical analysis

We calculated the proportion of the study sample that initiated and engaged with CDM care along with 95% confidence intervals. Descriptive statistics were used to characterize utilization of CDM services, substance abuse treatment (outpatient or inpatient treatment, excluding treatment for detoxification); and addiction pharmacotherapy. The latter two variables were assessed at a 3-month follow-up interview. Separate multivariable logistic regression models were used to identify predictors of 14-day initiation of CDM care and 30-day engagement with CDM care. All regression models were fit including a single independent variable of interest as well as a core set of covariates: age, sex, race/ethnicity, and study enrollment month. If more than one non-covariate independent variable was statistically significant ($P < 0.05$), then a single model was fit to include all statistically significant variables, again with core covariates. We verified that no pair of independent variables included in a regression model was highly correlated (i.e. >0.40), minimizing the potential for collinearity. Due to the exploratory nature of the analyses, no adjustments were made for multiple comparisons.

As secondary outcomes, we calculated the proportion of the study sample who met criteria for linkage with CDM care and continuation of CDM care using the Kaplan–Meier estimator to account for differential lengths of follow-up. To evaluate predictors of these CDM care utilization measures, we used a model-building approach similar to the one described above using Cox proportional hazards models. All statistically significant independent variables ($P < 0.05$) were combined into a single model with the same core covariates listed above. All analyses were completed using SAS/STAT software, Version 9.1, SAS Institute Inc. Cary, NC.

3. Results

3.1. Study subjects

This study's analytic sample was derived from participants enrolled in the AHEAD study randomized trial. Among 650 eligible individuals, 87% ($n = 563$) enrolled in the AHEAD randomized trial, and 282 were randomly assigned to have access to CDM addiction care comprising the sample for this study.

The baseline sociodemographic and health characteristics of the study sample ($n = 282$) are displayed in Table 1. Recruitment was largely from an inpatient detoxification unit (73%, 206/282). Most of the sample met criteria for both alcohol and drug dependence (56%), fewer (27%) for drug dependence only, and 17% for alcohol dependence only. Social problems, psychiatric and medical comorbidities were common including homelessness

Table 1
Characteristics of patients given access to chronic disease management (CDM) for substance dependence ($n = 282$).

	N (%)
Need variables	
Substance-related	
Substance dependence diagnosis[a]	
Alcohol only	49 (17%)
Drug only	76 (27%)
Alcohol and drug	157 (56%)
Overdose[b] (lifetime)	84 (30%)
Readiness to change[c]	
10	171 (61%)
Less than 10	111 (39%)
Social needs	
Homeless[d]	159 (56%)
Legal problems,[e] any	101 (36%)
Psychiatric needs	
Post traumatic stress disorder[f,g]	100 (36%)
Major depressive episode[f,h]	219 (78%)
Medical needs	
Comorbid medical condition[i]	133 (47%)
Substance use disorder -related medical condition[j]	189 (67%)
Enabling variables	
Unfavorable social network[k]	197 (70%)
Health insurance, any	221 (79%)
Mutual-help, 12-step program[l], any vs none (recent[m])	136 (48%)
Substance abuse treatment, residential[n] (recent[m])	60 (21%)
Psychiatric care[o], any (recent[m])	43 (15%)
Hospitalization, for medical problem (recent[m])	55 (20%)
Predisposing variables	
Age, mean (SD[p])	38.6 (9.9)
Female	84 (30%)
Race/ethnicity	
Black	93 (33%)
Hispanic	28 (10%)
Other	29 (10%)
White	132 (47%)
Study enrollment month[q]	
≥13	119 (42%)
7–12	65 (23%)
≤6	98 (35%)

[a] Composite International Diagnostic Interview Short Form (past year) and past 30 day drug use or heavy alcohol use.
[b] Overdose includes accidental and deliberate overdose of illegal drugs, over the counter medications, prescription medications, or alcohol.
[c] Readiness to change drinking or drug use: "How are ready are you to change your drinking or drug use?" using a scale from 1 to 10 with 10 indicating more readiness to change.
[d] Any shelter use or night on the street in the past 3 months.
[e] On probation, parole, pretrial release, or in diversion program (Drug Court).
[f] Mini International Neuropsychiatric Interview (MINI).
[g] Past month.
[h] Past 2 weeks.
[i] Katz Comorbidity Questionnaire.
[j] Includes seizures, heart failure, atrial fibrillation, rapid heat beat, hepatitis, cirrhosis, peripheral neuropathy, cancer of mouth/esophagus/stomach, skin infections, pneumonia, tuberculosis, gastritis, pancreatitis, anemia, septic arthritis, endocarditis, or blood clots.
[k] Environment favoring substance use defined as most or all of the people that you spend time with are either heavy/problem drinkers or heavy/problem drug users.
[l] Alcoholics Anonymous, Narcotic Anonymous, or Cocaine Anonymous.
[m] Past 3 months.
[n] Excludes admission for detoxification.
[o] Any counseling or therapy for emotional/psychological problems including full-day treatment, partial hospital program, or treatment by a psychiatrist.
[p] Standard deviation.
[q] Month of participant enrollment with respect to the first day of Addiction Health Evaluation and Disease management (AHEAD) cohort recruitment (September 11, 2006).

(56%), legal problems (36%), post-traumatic stress disorder (36%), major depression (78%), and substance-related medical conditions (67%).

Approximately half of the cohort (45%, 95% Confidence Interval [CI] 39–51%) met criteria for 14-day initiation and 23% (95% CI 18–28%) for 30-day engagement with CDM care (Table 2). By the end of study follow-up, more than three-fourths (81%, 95% CI 76–85%) of the cohort met criteria for linkage with CDM care and almost two-thirds (62%, 95% CI 56–68%) with continuation of CDM care.

Among those with CDM care linkage, the range of time from study enrollment to the second AHEAD visit was wide (1–458 days), however, the median was 12 days (interquartile range [IQR] 5, 34) and most (72% [164/227]) did so within 30 days of study entry. Similarly, among those with four or more AHEAD visits, the range of time from study enrollment to the fourth AHEAD visit was also remarkably wide (6–1059 days), but the median was 49 (IQR 21, 116) and most (67%) attended at least four AHEAD visits within 90 days.

Utilization of CDM care did not end with engagement. Participants who engaged with CDM care attended a median of 17 AHEAD visits (IQR 8, 27) over an extended period of time (median 514 days, IQR 180, 873). Participants with CDM continuation attended a median 14 visits (IQR 7, 25) over more than a year (median 550 days, IQR 287, 876).

We conducted a supplemental analysis to test whether engagement with CDM was associated with receipt of addiction treatments. Relative to those who did not engage with CDM care, a higher proportion of participants who engaged with CDM services utilized addiction treatment (79% vs 56%. respectively, P value = 0.001) and addiction pharmacotherapy (39% vs 18%, respectively, P value < 0.001).

3.2. Multivariable regression results

Major depressive episode was the only factor associated with initiation of CDM care (Table 3). Participants with major depressive episode had almost twice the odds of initiating CDM care (AOR 2.60, 95% CI 1.39, 4.87). Female sex was associated with lower odds of linkage with CDM care over the course of the study (Adjusted HR 0.67, 95% CI 0.49, 0.90).

Results of analyses (Tables 3 and 4) examining 30-day CDM engagement and CDM continuation were similar: younger age, female sex, and higher alcohol addiction severity were associated with lower odds of 30-day CDM engagement and a lower risk of CDM continuation. Later study enrollment was also associated with the latter CDM measure. Major depressive episode, while significant in the preliminary model, was not significant in a final model that combined all statistically significant factors and core covariates.

Since the sample was composed of individuals with alcohol dependence, drug dependence, or both alcohol and drug dependence, those with lower alcohol addiction severity may have been more likely to be primarily drug users who were seeking services available at the CDM clinic like office-based opioid therapy. We accounted for this possible confounding factor by adding a covariate for past 30-day opioid use, which did not substantially alter results.

4. Discussion

In this cohort of adults with substance dependence with access to CDM care located in primary care, approximately half initiated and a quarter engaged with CDM addiction care in the time frame specified by Washington Circle performance measures. Broadening the time interval, about three quarters of the sample made at least two CDM visits and about two-thirds at least four CDM visits over the course of the study. Regardless of the interval of time used to define these measures, most participants who engaged with CDM care continued to utilize CDM for more than a year.

An important point to consider is that recruitment of study participants was not based upon interest in utilizing CDM care. Although most of the study sample was recruited from a detoxification unit, they were not specifically seeking CDM care, or for that matter, any addiction treatment. Instead, they agreed to participate in a study that included an outpatient CDM clinic visit within primary care (i.e., the AHEAD clinic) or simply a referral to primary care. Therefore, participants in this study who were introduced to CDM care at study entry might return to utilize services in their "own time" instead of utilizing services within the timeframe specified in the WC performance measures. Accordingly, we found that most of the study sample returned to the clinic and CDM services during the follow-up period. Since engagement is considered to be a measure of patients' assessment of the appropriateness and attractiveness of treatment (McLellan et al., 2007), most participants appear to view CDM care favorably.

Because participants were not specifically seeking CDM care, initiation and engagement results approaching those of standard addiction treatment should be considered favorable. This study's initiation estimates are comparable to those reported by Massachusetts public sector addiction treatment sites and by the NCQA in an analysis of Medicaid data (42% for initiation and 27% for engagement) (Garnick et al., 2009). This study's engagement estimate was also comparable to Massachusetts public treatment sites but higher than the Medicaid rates reported by the NCQA (43% and 12%, respectively) (National Committee for Quality Assurance, 2010). Other treatment systems have reported higher treatment engagement rates (Kilbourne et al., 2006; Garnick et al., 2007).

A secondary objective was to evaluate predictors of initiation and engagement. Surprisingly, very few of the long list of variables we examined were significantly associated with initiation and engagement including factors associated with utilization of other types of addiction treatment, such as homelessness, legal problems, and readiness to change. Several of the variables associated with lower odds of ongoing CDM utilization have been noted with other types of addiction care. The association of younger age with lower likelihood to engage with CDM care is consistent with many others including one using WC performance measures (Garnick et al., 2007). As in the National Survey on Drug Use and Health, female gender has been associated with less addiction treatment (Wu et al., 2003). Since women living with children are less likely to use substance abuse treatment (Kertesz et al., 2006), women may have been unable to access ongoing addiction CDM due to family responsibilites, difficulty arranging transportation, and/or child care. Future CDM efforts may need to focus on women with substance dependence, possibly by providing child care, making home visits, or using alternate modalities to clinic visits like telephone contacts (McKay, 2009; Godley et al., 2010).

We also found that higher alcohol addiction severity was associated with lower odds of CDM engagement and continuation. The reasons for this are not clear. A few studies have found an association between higher alcohol addiction severity or higher frequency of alcohol use and less continuity of care after discharge from residential addiction treatment (Greenberg et al., 2002; Harris et al., 2006). As all participants in this study had substance dependence, those with lower alcohol severity may have been more likely to have comorbid opioid dependence; as those individuals would have been the comparison group for this characteristic, their desire to return to the clinic to access opioid pharmacotherapy may account for this finding. However, these results were similar even after

Table 2
Utilization of CDM care for substance dependence.

Outcome	Definition	Proportion (95% CI)
14-day initiation of CDM care[a]	≥2 CDM visits within 14 days after study entry	45% (39, 51)
Linkage with CDM care[b]	≥2 CDM visits between study entry and the end of the follow-up period[c]	81% (76, 85)
30-day engagement with CDM care[a]	≥2 CDM visits within 30 days after "14-day initiation" criteria are met	23% (18, 28)
Continuation of CDM care[b]	≥4 CDM visits between study entry and the end of the follow-up period[c]	62% (56, 68)

[a] Adapted from Washington Circle performance measures for initiation and engagement with outpatient addiction treatment.
[b] Kaplan–Meier survival estimate.
[c] Median follow-up was 550 days with interquartile range of 287–876.

Table 3
Factors associated with utilization of CDM care for substance dependence in multivariable regression models.

Need variables	14-day initiation[a]		CDM linkage[b]		30-day engagement[c]		CDM continuation[d]	
	OR[e]	95% CI	HR[f]	95% CI	OR[e]	95% CI	HR[f]	95% CI
Addiction-related needs								
Substance dependence diagnosis								
Drug only	1.47	0.62, 3.46	0.86	0.53, 1.37	2.43	0.85, 6.92	1.17	0.68, 2.01
Alcohol and drug	1.23	0.62, 2.41	0.87	0.60, 1.25	1.75	0.74, 4.13	0.96	0.62, 1.49
Alcohol only	1	–	1	–	1	–	1	–
Alcohol addiction severity[g]								
≥0.70	0.49	0.25, 0.95[h]	0.69	0.48, 1.00	**0.30**	**0.14, 0.67**	**0.49**	**0.32, 0.75**
0.30 to <0.70	0.55	0.28, 1.05	0.94	0.65, 1.35	**0.34**	**0.15, 0.74**	**0.68**	**0.45, 1.02**
<0.30	1	–	1	–	1	–	1	–
Drug addiction severity[g]								
≥0.38	1.31	0.68, 2.55	0.72	0.50, 1.03	1.47	0.66, 3.25	0.76	0.50, 1.15
0.25 to <0.38	1.06	0.57, 1.99	0.68	0.48, 0.96	1.24	0.59, 2.61	0.75	0.50, 1.13
<0.25	1	–	1	–	1	–	1	–
Overdose (lifetime) (yes vs no)	0.88	0.52, 1.51	0.91	0.68, 1.22	0.97	0.51, 1.85	0.88	0.63, 1.23
Readiness to change								
10	0.90	0.55, 1.48	0.89	0.67, 1.17	0.78	0.43, 1.41	0.82	0.59, 1.12
Less than 10	1	–	1	–	1	–	1	–
Social needs								
Homeless, any vs none	0.80	0.49, 1.30	0.86	0.66, 1.12	0.79	0.44, 1.42	0.78	0.58, 1.06
Legal problems, any vs none	1.59	0.96, 2.63	1.26	0.95, 1.66	1.37	0.75, 2.51	1.27	0.93, 1.73
Psychiatric needs								
Post traumatic stress disorder	1.11	0.67, 1.82	1.08	0.82, 1.42	1.11	0.61, 2.04	1.12	0.82, 1.53
Major depressive episode	**2.60**	**1.39, 4.87**	1.27	0.92, 1.76	**2.53**	**1.11, 5.80**	1.26	0.86, 1.84
Medical needs								
Comorbid medical condition	1.34	0.81, 2.20	1.27	0.96, 1.68	1.66	0.90, 3.07	1.11	0.81, 1.53
Substance use disorder-related medical condition	1.50	0.89, 2.53	1.13	0.85, 1.51	1.58	0.82, 3.04	1.18	0.84, 1.66
Enabling variables								
Unfavorable social network	1.17	0.68, 1.99	1.03	0.77, 1.37	1.32	0.68, 2.56	0.83	0.60, 1.16
Health insurance, yes vs no	0.66	0.35, 1.23	0.74	0.53, 1.05	1.27	0.57, 2.80	1.04	0.70, 1.54
AA, any vs none (recent)	1.08	0.67, 1.76	1.16	0.89, 1.51	1.10	0.61, 1.98	1.01	0.74, 1.37
Substance abuse treatment, residential (recent)	1.67	0.92, 3.04	1.20	0.86, 1.67	1.32	0.64, 2.72	1.36	0.94, 1.98
Psychiatric care, any (recent)	1.08	0.55, 2.08	1.09	0.76, 1.56	1.35	0.63, 2.90	1.17	0.76, 1.78
Hospitalization, for medical problem (recent)	1.76	0.96, 3.22	1.17	0.84, 1.61	1.42	0.71, 2.84	0.96	0.66, 1.41
Predisposing variables								
Younger age (1 SD decrease)	0.86	0.66, 1.13	0.90	0.78, 1.05	**0.66**	**0.47, 0.92**	0.84	0.71, 1.00
Female	0.65	0.38, 1.10	**0.67**	**0.49, 0.90**	**0.36**	**0.17, 0.75**	**0.61**	**0.43, 0.86**
Race/ethnicity								
Non-white vs white	1.07	0.63, 1.80	0.99	0.75, 1.32	0.86	0.46, 1.62	0.84	0.61, 1.17
Study enrollment month								
≥ 3	0.88	0.49, 1.58	0.92	0.67, 1.26	0.59	0.29, 1.21	**0.56**	**0.39, 0.82**
7–12	1.17	0.61, 2.23	1.04	0.73, 1.48	1.31	0.62, 2.77	**0.87**	**0.58, 1.28**
≤6	1	–	1	–	1	–	1	–

Bolded values indicate an association at $P < 0.05$.
[a] ≥2 CDM visits within 14 days of study entry.
[b] ≥2 CDM visits between study entry and the end of the follow-up period.
[c] ≥2 CDM visits within 30 days of achieving "14-day initiation".
[d] ≥4 CDM visits between study entry and the end of the follow-up period.
[e] Odds ratio from logistic regression models predicting 14-day initiation of CDM care and 30-day engagement with CDM care (separate models for each independent variable of interest, all adjusted for core covariates: age, gender, race/ethnicity, and study enrollment month).
[f] Hazard ratios from Cox proportional hazards models predicting CDM linkage and CDM continuation (separate models for each independent variable of interest, all adjusted for core covariates).
[g] Categories represent tertiles of Addiction Severity Index score (0-1) with 1 indicating higher severity.
[h] $P = 0.08$.

adjusting for past 30-day opioid use. Hence, these findings suggest that those with higher alcohol addiction severity may require additional effort (e.g., motivational enhancement therapy, contingency management, more desirable pharmacotherapies) to engage and continue with CDM care.

4.1. Limitations

The importance of these findings depends upon whether CDM care is found to be effective. There is growing evidence that integrated medical and addictions care is more effective than either

Table 4
Factors associated with "30-day engagement" and "CDM continuation" in the final multivariable model[a].

Independent Variable	30-day engagement		CDM continuation	
	OR	95% CI	HR	95% CI
Younger age (1 SD)	**0.64**	**0.47, 0.88**	**0.77**	**0.65, 0.93**
Female	**0.54**	**0.29, 0.98**	**0.60**	**0.43, 0.85**
Race/ethnicity (non-white vs white)	0.89	0.49, 1.59	0.93	0.67, 1.30
Alcohol Severity Index				
≥0.70	**0.34**	**0.17, 0.71**	**0.49**	**0.32, 0.75**
0.30 to <0.70	**0.48**	**0.24, 0.96**	**0.68**	**0.45, 1.02**
<0.30	1	–	1	–
Major depressive episode	1.42	0.73, 2.75	–	–
Study enrollment month				
≥13	0.89	0.45, 1.77	0.63	0.42, 0.94
7–12 months	1.31	0.65, 2.65	0.84	0.57, 1.25
≤6	1	–	1	–

[a] Results from one multivariable logistic regression model predicting "30-day engagement" with CDM care and one multivariable Cox proportional hazards model predicting "CDM continuation." Both models include all variables listed in the table.

alone (Weisner et al., 2001; Bartels et al., 2004; Saitz et al., 2005) and CDM care contains effective components of addiction treatment (e.g., case management) endorsed by the IOM and others. Even if CDM care per se is not found to be effective, these findings are of interest, given the growing interest in transforming a system of time-limited episodic addiction care to one that spans different stages of substance use recovery (McLellan et al., 2007) and even a lifetime.

Generalizability of the study's results is another consideration when interpreting these findings. Since CDM care is a type of care that likely does not exist in many places, the results of this study are not applicable to models of care now in widespread use. However, there are several reasons that these findings may be applicable to urban general healthcare settings where such clinics could be implemented. First, this study had broad eligibility criteria to allow individuals with significant social, psychiatric, and medical needs to participate in the study regardless of readiness to change or desire for specialty addiction treatment, and most who were eligible enrolled; these were individuals that one might find in general medical practice or populations often excluded from efficacy studies (Humphreys et al., 2008). Second, the AHEAD clinic was located within primary care but was not dependent upon addiction training of the primary medical care staff. Third, many services commonly needed by individuals with SD were subject to the usual financial, administrative, and limited availability constraints. These services included addiction pharmacotherapy, specialty addiction treatment, primary medical care, diagnostic testing, and all other medical treatment provided by the hospital. In sum, the CDM clinic itself had no barriers but for any services outside the clinic, there were constraints, and in those circumstances, CDM clinicians worked to facilitate receipt of those services. CDM is a treatment model that could be implemented in other sites, albeit with variation in local available specific treatments outside the clinic.

4.2. Implications

This study's findings are relevant to recent efforts to provide longitudinal rather than episodic addiction care to improve the quality of care (McLellan et al., 2005; Institute of Medicine, 2006; McKay, 2006; Dennis and Scott, 2007). Evidence is growing for the utility of continuing care interventions to maintain progress from an initial, more intensive treatment and facilitate earlier re-entry into addiction treatment with relapse (Patterson et al., 1997; McKay et al., 2005; Bennett et al., 2007). CDM care shares this longitudinal perspective of care and goes further by facilitating access to a range of addiction treatment modalities including addiction pharmacotherapy, engaging with patients to acquire self-management skills, and addressing comorbidities. CDM care for substance dependence co-located with primary care has the potential to address the needs of the overwhelming majority of individuals with SD who would otherwise not seek any addiction treatment (Saitz et al., 2008; McKay, 2009). Sustaining participation with CDM care is likely to be instrumental to its efficacy (McKay, 2005). Knowing more about factors associated with CDM initiation and engagement can help clinicians target those who are more likely to drop out of care.

In summary, patients with substance dependence appear to be willing to initiate and follow-up with CDM care, although often not in time frames specified by performance measures developed to evaluate the quality of outpatient addiction care. CDM has the potential for improving the quality of addiction care for people with addictions.

Role of funding source

Funding for this study was provided by the National Institute on Alcohol Abuse and Alcoholism (grant R01 AA010870) and the National Institute on Drug Abuse (grant R01 DA010019). This work was also supported, in part, by the Boston University Translational Science Institute from the National Center for Research Resources (UL1-RR025771). Dr. Samet receives support from NIAAA: K24 AA015674.

The sponsors had no further role in study design; in the collection, analysis, and/or interpretation of the data; in the writing of the report, or in the decision to submit the paper for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse, the National Institute on Alcohol Abuse and Alcoholism or the National Institutes of Health.

Contributors

Study design: R. Saitz, D.M. Cheng, and J.H. Samet; Protocol: R. Saitz, J. Witas, J.H. Samet; Literature search and summaries of previous work: T.W. Kim and J. Witas; Data management and statistical analysis: D.M. Cheng and M. Winter; First draft of the manuscript: T.W. Kim. All authors have contributed to and approved the final manuscript.

Conflict of interest

None.

References

Bartels, S.J., Coakley, E.H., Zubritsky, C., Ware, J.H., Miles, K.M., Arean, P.A., Chen, H., Oslin, D.W., Llorente, M.D., Costantino, G., Quijano, L., McIntyre, J.S., Linkins, K.W., Oxman, T.E., Maxwell, J., Levkoff, S.E., PRISM-E Investigators, 2004. Improving access to geriatric mental health services: a randomized trial comparing treatment engagement with integrated versus enhanced referral care for depression, anxiety, and at-risk alcohol use. Am. J. Psychiatry 161, 1455–1462.

Bennett, G.A., Roberts, H.A., Vaughan, T.E., Gibbins, J.A., Rouse, L., 2007. Evaluating a method of assessing competence in Motivational Interviewing: a study using simulated patients in the United Kingdom. Addict. Behav. 32, 69–79.

Cunningham, J.A., Blomqvist, J., 2006. Examining treatment use among alcohol-dependent individuals from a population perspective. Alcohol Alcohol. 41, 632–635.

De Alba, I., Samet, J.H., Saitz, R., 2004. Burden of medical illness in drug- and alcohol-dependent persons without primary care. Am. J. Addict. 13, 33–45.

Dennis, M., Scott, C.K., 2007. Managing addiction as a chronic condition. Addict. Sci. Clin. Pract. 4, 45–55.

Garnick, D.W., Horgan, C.M., Lee, M.T., Panas, L., Ritter, G.A., Davis, S., Leeper, T., Moore, R., Reynolds, M., 2007. Are Washington Circle performance measures associated with decreased criminal activity following treatment? J. Subst. Abuse Treat. 33, 341–352.

Garnick, D.W., Lee, M.T., Chalk, M., Gastfriend, D., Horgan, C.M., McCorry, F., McLellan, A.T., Merrick, E.L., 2002. Establishing the feasibility of performance measures for alcohol and other drugs. J. Subst. Abuse Treat. 23, 375–385.

Garnick, D.W., Lee, M.T., Horgan, C.M., Acevedo, A., Washington Circle Public Sector Workgroup, 2009. Adapting Washington Circle performance measures for public sector substance abuse treatment systems. J. Subst. Abuse Treat. 36, 265–277.

Gelberg, L., Andersen, R.M., Leake, B.D., 2000. The behavioral model for vulnerable populations: application to medical care use and outcomes for homeless people. Health Serv. Res. 34, 1273–1302.

Godley, M.D., Coleman-Cowger, V.H., Titus, J.C., Funk, R.R., Orndorff, M.G., 2010. A randomized controlled trial of telephone continuing care. J. Subst. Abuse Treat. 38, 74–82.

Greenberg, G.A., Rosenheck, R.A., Seibyl, C.L., 2002. Continuity of care and clinical effectiveness: outcomes following residential treatment for severe substance abuse. Med. Care 40, 246–259.

Harris, A.H., Humphreys, K., Bowe, T., Tiet, Q., Finney, J.W., 2010. Does meeting the HEDIS substance abuse treatment engagement criterion predict patient outcomes? J. Behav. Health Serv. Res. 37, 25–39.

Harris, A.H., McKellar, J.D., Moos, R.H., Schaefer, J.A., Cronkite, R.C., 2006. Predictors of engagement in continuing care following residential substance use disorder treatment. Drug Alcohol Depend. 84, 93–101.

Hasin, D.S., Stinson, F.S., Ogburn, E., Grant, B.F., 2007. Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Arch. Gen. Psychiatry 64, 830–842.

Humphreys, K., Harris, A.H., Weingardt, K.R., 2008. Subject eligibility criteria can substantially influence the results of alcohol-treatment outcome research. J. Stud. Alcohol Drugs 69, 757–764.

Institute of Medicine, 2006. Improving the Quality of Health Care for Mental and Substance-use Conditions: Quality Chasm Series. The National Academies Press, Washington DC.

Katz, J.N., Chang, L.C., Sangha, O., Fossel, A.H., Bates, D.W., 1996. Can comorbidity be measured by questionnaire rather than medical record review? Med. Care 34, 73–84.

Kertesz, S.G., Larson, M.J., Cheng, D.M., Tucker, J.A., Winter, M., Mullins, A., Saitz, R., Samet, J.H., 2006. Need and non-need factors associated with addiction treatment utilization in a cohort of homeless and housed urban poor. Med. Care 44, 225–233.

Kessler, R.C., Andrews, G., Mroczek, D., Ustun, B., Wittchen, H.U., 1998. The World Health Organization Composite International Diagnostic Interview Short Form (CIDI-SF). Int. J. Methods Psychiatric Res. 7, 171–185.

Kilbourne, A.M., Salloum, I., Dausey, D., Cornelius, J.R., Conigliaro, J., Xu, X., Pincus, H.A., 2006. Quality of care for substance use disorders in patients with serious mental illness. J. Subst. Abuse Treat. 30, 73–77.

McKay, J.R., 2005. Is there a case for extended interventions for alcohol and drug use disorders? Addiction 100, 1594–1610.

McKay, J.R., 2006. Continuing care in the treatment of addictive disorders. Curr. Psychiatry Rep. 8, 355–362.

McKay, J.R., 2009. Continuing care research: what we have learned and where we are going. J. Subst. Abuse Treat. 36, 131–145.

McKay, J.R., Lynch, K.G., Shepard, D.S., Pettinati, H.M., 2005. The effectiveness of telephone-based continuing care for alcohol and cocaine dependence: 24-month outcomes. Arch. Gen. Psychiatry 62, 199–207.

McLellan, A.T., 2002. Have we evaluated addiction treatment correctly? Implications from a chronic care perspective. Addiction 97, 249–252.

McLellan, A.T., Chalk, M., Bartlett, J., 2007. Outcomes, performance, and quality: what's the difference? J. Subst. Abuse Treat. 32, 331–340.

McLellan, A.T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., Argeriou, M., 1992. The fifth edition of the addiction severity index. J. Subst. Abuse Treat. 9, 199–213.

McLellan, A.T., Lewis, D.C., O'Brien, C.P., Kleber, H.D., 2000. Drug dependence, a chronic medical illness: implications for treatment, insurance, and outcomes evaluation. JAMA 284, 1689–1695.

McLellan, A.T., McKay, J.R., Forman, R., Cacciola, J., Kemp, J., 2005. Reconsidering the evaluation of addiction treatment: from retrospective follow-up to concurrent recovery monitoring. Addiction 100, 447–458.

National Committee for Quality Assurance, 2010. The State of Health Care Quality 2009. National Committee for Quality Assurance, USA.

National Quality Forum, 2007. National Voluntary Consensus Standards for the Treatment of Substance use Conditions: Evidence-Based Treatment Practices: A Consensus Report. National Quality Forum, Washington DC.

Patterson, D.G., Macpherson, J., Brady, N.M., 1997. Community psychiatric nurse aftercare for alcoholics: a five-year follow-up study. Addiction 92, 459–468.

Rea, H., McAuley, S., Stewart, A., Lamont, C., Roseman, P., Didsbury, P., 2004. A chronic disease management programme can reduce days in hospital for patients with chronic obstructive pulmonary disease. Int. Med. J. 34, 608–614.

Roy-Byrne, P.P., Katon, W., Cowley, D.S., Russo, J., 2001. A randomized effectiveness trial of collaborative care for patients with panic disorder in primary care. Arch. Gen. Psychiatry 58, 869–876.

Saitz, R., Horton, N.J., Larson, M.J., Winter, M., Samet, J.H., 2005. Primary medical care and reductions in addiction severity: a prospective cohort study. Addiction 100, 70–78.

Saitz, R., Larson, M.J., Labelle, C., Richardson, J., Samet, J.H., 2008. The case for chronic disease management for addiction. J. Addict. Med. 2, 55–65.

Scott, C.K., Dennis, M.L., 2009. Results from two randomized clinical trials evaluating the impact of quarterly recovery management checkups with adult chronic substance users. Addiction 104, 959–971.

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry 59 (Suppl. 20), 22–33, quiz 34-57.

Simon, G.E., Katon, W.J., VonKorff, M., Unutzer, J., Lin, E.H., Walker, E.A., Bush, T., Rutter, C., Ludman, E., 2001. Cost-effectiveness of a collaborative care program for primary care patients with persistent depression. Am. J. Psychiatry 158, 1638–1644.

Smith, K.L., Horton, N.J., Saitz, R., Samet, J.H., 2006. The use of the mini-mental state examination in recruitment for substance abuse research studies. Drug Alcohol Depend. 82, 231–237.

Substance Abuse and Mental Health Services Administration, Office of Applied Studies, 2008. Treatment Episode Data Set (TEDS): 2005. Discharges from Substance Abuse Treatment Services. DHHS, Rockville, MD.

Tucker, J.A., Foushee, H.R., Simpson, C.A., 2009. Increasing the appeal and utilization of services for alcohol and drug problems: what consumers and their social networks prefer. Int. J. Drug Policy 20, 76–84.

Wagner, E.H., 2000. The role of patient care teams in chronic disease management. BMJ 320, 569–572.

Weisner, C., Mertens, J., Parthasarathy, S., Moore, C., Lu, Y., 2001. Integrating primary medical care with addiction treatment: a randomized controlled trial. JAMA 286, 1715–1723.

Whellan, D.J., Hasselblad, V., Peterson, E., O'Connor, C.M., Schulman, K.A., 2005. Metaanalysis and review of heart failure disease management randomized controlled clinical trials. Am. Heart J. 149, 722–729.

Wu, L.T., Ringwalt, C.L., Williams, C.E., 2003. Use of substance abuse treatment services by persons with mental health and substance use problems. Psychiatr. Serv. 54, 363–369.

Statement of the American Society of Addiction Medicine Consensus Panel on the Use of Buprenorphine in Office-Based Treatment of Opioid Addiction

Mark L. Kraus, MD, FASAM, Daniel P. Alford, MD, MPH, FACP, Margaret M. Kotz, DO, Petros Levounis, MD, FASAM, Todd W. Mandell, MD, Marjorie Meyer, MD, Edwin A. Salsitz, MD, FASAM, Norman Wetterau, MD, FASAM, and Stephen A. Wyatt, DO

Objectives: Opioid addiction affects over 2 million patients in the United States. The advent of buprenorphine and the passage of the Drug Addiction Treatment Act in 2000 have revolutionized the opioid treatment delivery system by granting physicians the ability to administer office-based opioid treatment (OBOT), thereby giving patients greater access to treatment. The purpose of this consensus panel was to synthesize the most current evidence on the use of buprenorphine in the office-based setting and to make recommendations that will enable and allow additional physicians to begin to treat opioid-addicted individuals.

Methods: Literature published from 2000 to 2009 was searched using the PubMed search engine and yielded over 375 articles published in peer-reviewed journals, including some that were published guidelines. These articles were submitted to a consensus panel composed of researchers, educators, and clinicians who are leaders in the field of addiction medicine with specific expertise in the use of OBOT. The panel discussed results and agreed upon consensus recommendations for several facets of OBOT.

Results: On the basis of the literature review and consensus discussions, the panel developed a series of findings, conclusions, and recommendations regarding the use of buprenorphine in office-based treatment of opioid addiction.

Conclusions: Therapeutic outcomes for patients who self-select office-based treatment with buprenorphine are essentially comparable to those seen in patients treated with methadone programs. There are few absolute contraindications to the use of buprenorphine, although the experience and skill levels of treating physicians can vary considerably, as can access to the resources needed to treat comorbid medical or psychiatric conditions—all of which affect outcomes. It is important to conduct a targeted assessment of every patient to confirm that the provider has resources available to meet the patient's needs. Patients should be assessed for a broad array of biopsychosocial needs in addition to opioid use and addiction, and should be treated, referred, or both for help in meeting all their care needs, including medical care, psychiatric care, and social assistance. Current literature demonstrates promising efficacy of buprenorphine, though further research will continue to demonstrate its effectiveness for special populations, such as adolescents, pregnant women, and other vulnerable populations. Since the time of this review, several new studies have provided new data to continue to improve our understanding of the safety and efficacy of buprenorphine for special patient populations.

Key Words: buprenorphine, office-based treatment, opioid addiction

(*J Addict Med* 2011;5: 254–263)

From the Yale University School of Medicine (MLK), New Haven; Primary Care, General and Internal Medicine/Addiction Medicine (MLK), Westside Medical Group, PC, Waterbury; Chief Medical Officer (MLK), Connecticut Counseling Centers, Danbury, CT; Boston University School of Medicine (DPA); Office-Based Opioid Treatment and MASBIRT Programs (DPA), Boston Medical Center, Boston, MA; Case Western Reserve University School of Medicine (MMK); University Hospitals of Cleveland (MMK), Cleveland, OH; The Addiction Institute of New York (PL); Division of Addiction Psychiatry (PL), St Luke's & Roosevelt Hospitals; Columbia University College of Physicians & Surgeons (PL), New York, NY; Community Substance Abuse Centers, Inc (TWM), Westfield, MA; Opiate Treatment Program During Pregnancy (MM); Ob/Gyn (MM), University of Vermont Medical Center, Burlington, VT; Office-Based Opioid Therapy (EAS), Department of Medicine, Beth Israel Medical Center, New York, NY; Private Practice of Family Medicine (NW), Tri-County Family Medicine, Dansville, NY; and Dual Diagnosis Program (SAW), Middlesex Hospital, Middletown, CT.

Received for publication February 9, 2011; accepted August 3, 2011.

Supported by an unrestricted educational grant to ASAM from Reckitt-Benckiser Pharmaceutics Inc, manufacturer of Subutex and Suboxone.

The work reported in this article was conducted by a consensus panel and a national peer review panel appointed by the board of directors of the American Society of Addiction Medicine.

The authors are solely responsible for the content of this article and the decision to submit it for publication. No statement in this article should be construed as an official position of ASAM.

The following authors have declared that they have no known or potential conflicts of interest: Drs Alford, Kotz, Mandell, Meyer, Sullivan, Wetterau, and Wyatt. Dr Kraus has disclosed that he is a member of advisory panels to Reckitt-Benckiser, Forrest, and Pfizer. Dr Levounis has disclosed that he is a member of advisory panels to Astra Zeneca, Forest, and Pfizer. Dr Salsitz has disclosed that he is a member of advisory panels to Reckitt-Benckiser, Pfizer, and PriCara.

Send correspondence to Mark L. Kraus, MD, FASAM, 714 Chase Parkway, Waterbury, CT 06708. E-mail: mlkwmg@snet. And send reprint requests to: Angela Warner, ASAM, 4601 North Park Ave, Suite 101 Upper Arcade, Chevy Chase, MD 20815. E-mail: awarner@asam.org.

ISSN: 1932-0620/11/0504-0254

DOI: 10.1097/ADM.0b013e3182312983

Since 2002, a large body of evidence has become available, reflecting the experience of US researchers and

clinicians. This evidence is reflected in articles published in peer-reviewed journals, as well as in guidelines issued by various organizations and agencies (Health Resources and Services Administration, 2004; McNicholas, 2004; Wedam et al., 2007; Center for Substance Abuse Treatment, 2006a; Center for Substance Abuse Treatment, 2006b; Gordon and Krumm, 2008; Baxter, 2009).

To make this information more readily available to practicing physicians, as well as to encourage additional physicians to begin treating opioid-addicted persons, members of the American Society of Addiction Medicine consensus panel engaged in a critical examination of the scientific literature and employed their considerable clinical experience in reaching consensus as to recommended patient care practices. In doing so, panel members recognized that advice is not an adequate substitute for the knowledge and skills of practicing physicians who are engaged in developing treatment regimens tailored to the needs of individual patients.

The panel also recognized that not all treatment providers would be able to conform to each of the strategies recommended here. Instead, providers are encouraged to consider the panel's findings, conclusions, and recommendations in the context of the individual patient and their overall practice. Information and recommendations provided in this document are not intended to create a legal standard of care for any physician or to interfere with his or her clinical judgment or practice of medicine.

METHODS

Literature published from 2000 to 2009 was the subject of a PubMed search. The search yielded 376 articles published in peer-reviewed journals. Consensus reports from the federal Center for Substance Abuse Treatment and other authoritative sources also were included in the review.

Articles and published guidelines were submitted to a consensus panel composed of researchers, educators, and clinicians who have expertise in the use of buprenorphine. On the basis of the literature review and consensus discussions, the panel developed a series of findings, conclusions, and recommendations regarding the use of buprenorphine in office-based treatment of opioid addiction. Members agreed on the evidence for buprenorphine's overall efficacy and safety, as well as contraindications to its use.

Multiple drafts of the consensus panel's work were submitted to a national peer review panel, whose members were asked to evaluate the documents for scientific accuracy and clinical relevance. That work is presented here.

RESULTS

Patient Management With Buprenorphine

Induction

Patients who are currently physically dependent on opioids should be in moderate opioid withdrawal before the first buprenorphine induction dose. Patients are instructed to stop taking their opioid, and wait until they develop moderate spontaneous withdrawal. If a patient is not in withdrawal, and is given buprenorphine, precipitated withdrawal may occur. The Clinical Opioid Withdrawal Scale is a useful and validated assessment tool. The initial dose is 2 to 4 mg, and the total first day dose is up to 12 to 16 mg (Johnson et al., 2003; McNicholas, 2004; Batki, 2005; Marsch et al., 2006; Stephen, 2006; Baxter, 2009). During induction, patients should be frequently assessed for signs of overmedication. There is no data as to the specific time interval during which overmedication should be assessed. Therefore, it should be approached based on an individual patient basis.

Patients requesting transfer from methadone to buprenorphine should gradually taper their methadone dose to 30 to 40 mg and remain clinically stable on that dose before starting buprenorphine induction. Because methadone has a long and variable half-life, patients will need to discontinue methadone for at least 36 hours and often up to 72 hours to experience moderate withdrawal before proceeding with buprenorphine induction (McNicholas, 2008).

Patients should be advised to avoid driving or operating other machinery until their dose is stabilized and they are familiar with the effects of buprenorphine.

During induction and stabilization, patients should be assessed frequently for signs of overmedication or undermedication, and dose adjustments should be made accordingly (Johnson, 2003; McNicholas, 2004; Batki, 2005).

Consensus of the Panel

The buprenorphine/naloxone combination product should be used for induction as well as for stabilization and maintenance. The exception is pregnant women who are candidates for buprenorphine treatment, who should be inducted and maintained on the buprenorphine monoproduct (see the discussion of pregnancy).

In opioid dependent patients undergoing induction who exhibit signs of precipitated withdrawal, the physician has 2 options:

1. Continue with buprenorphine induction by continuing to give additional doses of buprenorphine up to 16 mg or until signs and symptoms of withdrawal abate;
2. Or to stop induction when the patient exhibits withdrawal symptoms, treat withdrawal symptomatically (eg, clonidine, antidiarrheals, nonsteroidal anti-inflammatory drugs) and instruct the patient to continue to abstain from opioids and return the following day for reassessment of induction.

The timing of buprenorphine induction requires care to avoid overdose (eg, in a patient who has been using central nervous system depressants such as alcohol or benzodiazepines in addition to opioids) or underdose (eg, triggering a re-emergence of opioid craving).

Stabilization

The stabilization phase is focused on finding the optimal dose for the individual patient. This dose should eliminate all withdrawal symptoms, decrease opioid craving, eliminate other opioid use, and provide maximal functional status (Joseph et al., 2000; Baxter, 2009).

Most patients stabilize on 8 to 24 mg/day (Comer et al., 2005a; Comer et al., 2005b). Rarely, there is a need to go up to 32 mg for the highly tolerant patient. The primary concern

in going to these larger doses is the much greater potential for diversion.

Certain medical factors may cause a patient's dosing requirements to change. These include (but are not limited to) starting, stopping, or changing the dose of other prescription medications; onset and progression of pregnancy; onset of menopause; progression of liver disease; and significant increase or decrease in weight (Baxter, 2009).

Relapse should always be ruled out as a reason for loss of stability. Continued or resumed use of short-acting opioids during treatment with buprenorphine may increase tolerance and render the buprenorphine dose inadequate (Stephen, 2008). If a short-acting opioid of abuse produces euphoria, the buprenorphine dose may be increased to block this effect. A dose increase also may help to suppress drug cravings (Leavitt et al., 2000). Ideally, receipt of opioids from multiple providers should be avoided. However, in cases where this is not so, coordination with other prescribing physicians to limit the number of short-acting opioids obtained by prescription is essential (Baxter, 2009).

Consensus of the Panel

There is no precise way to determine in advance the optimal dose for a particular patient. Because buprenorphine has a long plasma half-life and an even longer duration of action at the mu opioid receptor, 5 days should be allowed between dose adjustments to assess the effect. While most patients stabilize on a dose of 8 to 24 mg/day as demonstrated by the data, many will not need a dose higher than 16 mg/day. This is further supported by Comer et al. in their study noting that brain mu receptors are approximately 90% saturated at a dose of 16 mg/day demonstrated on neuroimaging.

Detoxification/Medically Supervised Withdrawal Management

Few studies have evaluated predictors, mediators, and moderators of treatment success for medically supervised withdrawal from opioids conducted in outpatient settings.

Consensus of the Panel

Detoxification using buprenorphine is not technically difficult, but long-term abstinence following such detoxification appears as difficult to achieve as with other medications. Arguably, detoxification is best conceptualized not as definitive treatment, but as a preparatory and stabilizing introduction to other forms of care.

The most effective withdrawal method involves stabilizing the opioid dependent patient with buprenorphine and then tapering the dose over time by 2 mg every 5 days until the taper is completed. Evidence comparing buprenorphine with methadone is limited, but it appears that completion of withdrawal may be more likely with buprenorphine and withdrawal symptoms may resolve more quickly with buprenorphine than with methadone (Gwoing et al., 2006; Gowing et al., 2009).

Maintenance Treatment

Except in patients whose addictive disorders are of brief duration, the best outcomes occur with long-term medication maintenance with methadone or buprenorphine accompanied by appropriate psychosocial interventions (Collins and McAllister, 2007; Kleber, 2007; Soeffing et al., 2009; Stotts et al., 2009; World Health Organization, 2009). The optimum duration of maintenance is unclear, but may involve long-term or even lifetime medication use (Kleber, 2007; World Health Organization, 2009). This is similar to the treatment of other chronic diseases, such as hypertension, diabetes, or asthma. In the maintenance treatment paradigm, the goal is not to "get off" the buprenorphine, but rather to achieve maximal function both at home and at work.

Generally 8 to 24 mg/day of buprenorphine will be an adequate maintenance dose. Some patients may require a higher dose up to 32 mg or a lower dose as noted above. In positron emission tomographic scan studies, approximately, 92% of mu opioid receptors were occupied by buprenorphine at a dose of 16 mg/day (Comer et al., 2005a; Comer et al., 2005b). It is unclear how positron emission tomographic scan images translate into clinical outcomes such as withdrawal, craving, and treatment retention (Johnson et al., 2003; Sporer, 2004). Doses higher than 24 mg should prompt a thorough review of the patient's rehabilitation status. To evaluate patient progress and success of maintenance therapy, physicians should assess patients regularly for relapse and instability. In managing these challenges, some of the consensus panel recommendations include:

- Increasing frequency of visits
- Adding additional psychosocial interventions
- Increasing drug dose (if not higher than the maximum daily dose already)
- Decrease prescription interval
- Increase level of care
- Initiate a unilateral involuntary taper toward medication discontinuation
- Consider switching medication management to methadone

Consensus of the Panel

As Kleber (Kleber, 2007) has noted, medications are available to treat opioid addiction although none are curative. Medications can, however, markedly diminish withdrawal symptoms and craving, and block opioid euphoric effects if patients relapse, and enhance the efficacy of psychosocial interventions.

Relapse Prevention

Among the major challenges confronting patients in treatment is the prevention of relapse, which is a risk even with successful treatment interventions (White, 2007).

Specific precipitants of relapse vary substantially from one experience to the next, even in the same individual (Connors et al., 1996). Attributing causality is even more complex in patients who have co-occurring medical or psychiatric disorders. In a survey, Daley and colleagues identified factors that contributed to relapse including inability to manage stress or negative emotional states (69%); interpersonal conflicts with family or others (29%); poor adherence to the treatment regimen (25%); negative thinking (11%); and insufficient motivation to change (10%) (Daley et al., 1998).

Consensus of the Panel

The following principles, which are common to many models of relapse prevention (Marlatt and Gordon, 1985; Tims and Leukefeld, 1987; Dimeff and Marlatt, 1995; Amato et al., 2008a; Amato et al., 2008b), can minimize the risk of relapse and attenuate the severity of a relapse episode:

1. Identifying environmental cues and stressors that act as relapse triggers.
2. Learning to identify and manage negative emotional states.
3. Working toward a more balanced lifestyle.
4. Developing skills to cope with stressful life events.
5. Understanding and managing craving.
6. Learning to identify and interrupt lapses and relapses.
7. Developing a recovery support network, such as joining a self-help group.
8. Utilizing clinical resources available to patients, such as counseling.

PATIENT SELECTION

Patient Assessment

The assessment should include:

1. Establishment of the diagnosis (such as, the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition) of opioid dependence including the duration and severity of the illness.
2. Discussion of current opioid use history in terms of when, what and how much opioid the patient most recently used.
3. Documentation of the patient's substance use history, including alcohol and other drugs of abuse.
4. Identification and referral of patients who need medically supervised withdrawal management from alcohol, benzodiazepines, or other sedatives.
5. Identification of comorbid medical and psychiatric conditions and disorders and to determine how, when and where they will be addressed.
6. Screening for communicable diseases and address them as needed.
7. Assessment of the patient's access to social supports, family, friends, employment, housing, finances, and legal problems.
8. Evaluation of the patient's readiness to participate in treatment.

Laboratory tests should include; liver function tests, human immunodeficiency virus (HIV) and viral hepatitis serologies, pregnancy test for women and urine toxicology screening for naturally occurring opioids (eg, such as heroin which is detected as morphine), synthetic and semisynthetic opioids (methadone, oxycodone), and other commonly abused drugs such as cocaine, amphetamines, and benzodiazepines (Gordon and Krumm, 2008).

Consensus of the Panel

Consensus opinion is that an initial patient assessment is of higher quality when it includes a medical and psychiatric history, a substance abuse history, and an evaluation of family and psychosocial supports, as well as pregnancy testing for all women of childbearing age. The physical examination should be focused on evaluating neurocognitive function and identification of sequelae of opioid addiction or severe hepatic dysfunction (Gordon and Krumm, 2008). The decision to initiate medication-assisted therapy begins with an evaluation of the patient to confirm the diagnosis of opioid dependence. An ideal candidate for office-based treatment with buprenorphine is an individual who will comply with each issue noted in the patient consent form such as withdrawal symptoms, risks of combining buprenorphine with other medications, directions for taking medication, cost of medication, and potential side effects (McNicholas, 2004).

Relative Contraindications to the Use of Buprenorphine

Concurrent Use of Sedative-Hypnotics (Including Alcohol)

Individuals with current, active alcohol dependence rarely are appropriate candidates for office-based treatment with buprenorphine (Fishman et al., 2005; McNicholas, 2008).

The combination of buprenorphine with sedative-hypnotic medications has been associated with deaths (Lavie et al., 2005). If concomitant treatment is deemed necessary, the doses of both medications may need to be reduced.

Elevated liver function tests: patients with elevated liver function 3 to 5 times above normal should not be considered for buprenorphine.

Consensus of the Panel

Patients who are dependent or abusing sedative hypnotics, alcohol, or both are rarely appropriate for office-based opioid treatment (OBOT) with buprenorphine. These patients should undergo careful clinical evaluation and should be considered for OBOT only if all of the following apply: clinical indication; willingness to discontinue sedative hypnotics, alcohol, or both by undergoing medically supervised withdrawal; and success in discontinuing hypnotics, alcohol, or both. In addition, patients with elevated liver functions tests 3 to 5 times greater than normal should not be considered for treatment with buprenorphine. These patients may be considered if they are willing to and have successfully discontinued sedative hypnotics, typically through medically supervised withdrawal.

Adolescents

Few studies have systematically evaluated buprenorphine in the treatment of adolescents, although there is good evidence that patients younger than 18 years are at particularly high risk for serious complications of addiction, including overdose deaths, suicide, HIV and other infectious diseases (Levy et al., 2007; Fiellin, 2008). Woody and colleagues (Woody et al., 2008) conducted 12-week clinical trials at 6 community programs for patients aged 15 to 21 years who were randomized to either 12 weeks of buprenorphine/naloxone treatment or a 14-day taper. Adolescents in the treatment group remained in counseling/ancillary treatment longer than those that were rapidly tapered (70 vs 20.5%, $P = 0.001$). Woody and colleagues concluded that continuing treatment with buprenorphine/naloxone improved outcomes compared with short-term detoxification, although further research is needed to assess the

efficacy and safety of longer-term treatment with buprenorphine for adolescents.

Consensus of the Panel

Buprenorphine/naloxone may be considered in adolescents for whom the balance of risks/benefits is considered favorable, considering such factors as: severity of addiction, previous failure, or low likelihood of success of other treatment approaches and overall risk of relapse. Furthermore, risks and benefits of using buprenorphine in adolescents should be discussed between providers and patients (and parents or guardians if patient is less than 18 years of age) on an individual basis.

Pregnant Women

Until recently, in the United States, methadone was the standard of care for pregnant women addicted to opioids. Since this review was conducted, however, new research has demonstrated promising safety and efficacy data for use of buprenorphine in pregnant women (Jones et al., 2009). Pregnant women should be offered either methadone or buprenorphine. Methadone has been shown to be safe and effective for both the pregnant woman and the neonate (Anderson and Kearney 2000; Jones et al., 2005; Vavrinková and Binder 2007). Buprenorphine (as is methadone) has been labeled as a Category C because there was insufficient evidence to establish its safety during pregnancy. Since the release of the 2009 study, buprenorphine monotherapy is a reasonable choice and appears to be as safe as methadone in pregnancy.

Buprenorphine maintained pregnancies also suggest that there is a lower severity of neonatal abstinence syndrome. Several studies have recently been published that have demonstrated that buprenorphine offers a substantial efficacy advantage over the current standard of care with oral morphine (Kraft et al., 2011; Unger et al., 2011). Other potential treatment of neonatal abstinence syndrome is methadone (Bio et al., 2011). This consensus panel review did not evaluate alternative delivery forms other than sublingual tablets (Note: more information and data maybe found in the references cited).

Breast-feeding

The safe use of buprenorphine during breast-feeding is not clearly delineated. However, the benefits of breast-feeding are multiple, including a natural strengthening of the maternal-child bond, which is of particular importance for this patient population. Further research will continue to clarify details regarding the use of buprenorphine in breast-feeding but until then the risks and benefits should be discussed and balanced on an individual patient basis (Lejeune et al., 2005; Briggs et al., 2008).

Consensus of the Panel

Short-term data on pregnancy and neonatal outcomes at the time of this review may indicate buprenorphine monotherapy for treatment of opioid dependent pregnant women is safe. Studies released since this review demonstrate promising safety and efficacy for the use of buprenorphine in pregnant women.

Although the available data are insufficient to firmly establish the safety of breast-feeding for mothers maintained on buprenorphine, the low theoretical risk should be balanced against the well-documented benefits of breast-feeding to both mother and neonate.

Pregnant patients require extensive counseling and community resources for recovery and parenting success. Integration of services and communication among all providers is essential for office-based treatment. The buprenorphine provider should work with the obstetric and pediatric providers to plan all aspects of care within the community. Since these consensus statements were written, additional research has been published that has demonstrated and supported the safety of buprenorphine in pregnant patients. Initial outcomes from these studies are positive demonstrating good outcomes for both mothers and neonates. A full discussion of this research is beyond the current scope of these guidelines, but readers should refer to the referenced literature to obtain further details (Jones et al., 2010).

Patients with Acute and Chronic Pain

In the United States, the parenteral formulation of buprenorphine is approved by the Food and Drug Administration for pain but not addiction treatment, while the sublingual formulation is approved for addiction but not pain treatment. Small studies in Europe and Asia demonstrate analgesic efficacy of the sublingual formulation (0.2-0.8 mg q 6-8 h) in opioid naïve postoperative pain (Edge et al., 1979; Moa and Zetterstrom, 1990). Parenteral analgesic potency is about 30 times that of morphine.

Consensus of the Panel

Several possible approaches exist for treating acute pain requiring opioid analgesia in the patient on buprenorphine therapy. With such limited clinical experience, mutiple treatment approaches based on pharmacologic principles have been published (Alford et al., 2006). The most effective approach will be elucidated with increased clinical experience. Currently there are insufficient data to recommend sublingual buprenorphine for the treatment of acute or chronic pain in patients with a history of opioid dependence.

Patients with HIV Disease

Buprenorphine should be used cautiously in combination with HIV antiretroviral medications that may inhibit, induce, or be metabolized by the cytochrome P450 3A4 enzyme system. Protease inhibitors inhibit cytochrome P450 3A4. Metabolism of buprenorphine, the antiretroviral medications, or both may be altered when they are combined. In some cases, therapeutic blood levels of antiretrovirals may need to be monitored (McCance-Katz, 2005; McCance-Katz et al., 2006a; McCance-Katz et al., 2006b).

Consensus of the Panel

While buprenorphine may be effectively used to treat patients with HIV, its use should be with caution concerning possible but as yet not clinically relevant drug-drug interactions. While drug/drug interactions remain a consideration with buprenorphine, they are likely less of a concern than when treating patients with methadone. Furthermore, successful use

of buprenorphine to treat HIV-infected, opioid-addicted patients has been demonstrated in multiple studies (Moatti et al., 2000; Berson et al., 2001; McCance-Katz, 2005).

Patients With Hepatitis and Other Liver Disorders

Viral hepatitis (especially infection with hepatitis B virus or hepatitis C virus) is common among individuals with up to 60% to 90% of injection drug users being infected with hepatitis C (Berson et al., 2001; Cazorla et al., 2005; Bruce and Altice, 2006). Therefore, patients with viral hepatitis who have opioid dependence and should be evaluated and treated appropriately (Backmund et al., 2001).

Consensus of the Panel

Buprenorphine treatment is not contraindicated by mildly elevated liver enzymes, although liver enzyme levels should be monitored. The threshold for elevated liver function for starting or discontinuing buprenorphine therapy is an elevated liver enzyme level of 3 times above normal. Patients with a history of injection drug use should be strongly encouraged to undergo immunization for hepatitis A and B, taking into account individual patient factors and appropriateness for vaccination.

Patients With Other Medical Conditions and Complications

Buprenorphine may be superior to methadone for the treatment of opioid dependence for patients with underlying cardiopulmonary disease or at risk for respiratory compromise, as it is less likely to cause respiratory depression (Gordon and Krumm, 2008).

Consensus of the Panel

Medical comorbidities may complicate the treatment of opioid addiction. Buprenorphine may be preferrable to methadone for specific medical conditions, though the evidence does not necessarily support which specific conditions.

Patients With Psychiatric Comorbidities

Coexisting psychiatric disorders are present in 20% to 60% of the persons entering addiction treatment, especially older individuals, those living in urban areas, patients who are incarcerated, or patients of a lower socioeconomic status (Robins et al., 1991; Kessler et al., 1994; Room, 1998; Sacks and Ries, 2005).

Patients with co-occurring psychiatric disorders have more difficulty engaging in, participating in, and completing addiction treatment, and generally have poorer prognoses than patients with diagnoses of either substance use or mental disorder alone (Kessler et al., 1994; Dausey and Desai, 2003). Untreated or inadequately treated psychiatric disorders can interfere with the effective treatment of addiction (Ziedonis et al., 2003; Khalsa et al., 2008. Patients with major depression or dysthymia are more likely to use illicit drugs during treatment than patients who do not suffer from depression (Sacks and Ries, 2005).

The presence of comorbid psychiatric disorders should not exclude patients from admission to office-based treatment with buprenorphine if outpatient treatment of both diseases can be accomplished (Sacks and Ries, 2005).

Consensus of the Panel

It is important to determine whether psychiatric symptoms are independent of the substance use or are substance-induced as this may inform treatment approach. Regardless, all patients with psychiatric symptoms should be evaluated and adequately treated. In the latter case, stability in the addiction treatment regimen should be the first therapeutic step (Ziedonis et al., 2003).

However, in patients with very severe psychiatric disease, the reverse treatment sequence may be more reasonable. In these patients, treatment using maintenance buprenorphine should be considered following stabilization of illness.

DISCUSSION

Although almost 2 million persons in the United States abuse or are addicted to opioids—prescription and illicit-–recent data suggest that nearly 80% do not receive treatment for their disorder (Kleber, 2007; Becker et al., 2008; Tetrault et al., 2008).

The use of buprenorphine and buprenorphine/naloxone combination in office-based primary care has improved access to care. Multiple studies have shown that buprenorphine treatment of addiction can be successfully integrated into office practice by physicians who are not addiction specialists (Fiellin et al., 2008).

In most cases, treatment is required for a long period or even throughout life (Kleber 2007; World Health Organization, 2009). Such long-term care, which is common to many medical conditions, should not be seen as a failure of treatment but as a cost-effective way to prolong life and improve quality of life by supporting the natural and long-term process of change and recovery. While the consensus panel did not make a formal recommendation on the frequency of monitoring and the use of varying methods for monitoring patient progress, we agreed that this is often dependent on physician preference as well as the individual patient. Therefore, urine toxicology, prescription supply/interval, etc, may be considered for each individual patient case.

Recent studies indicate that buprenorphine can be used safely and effectively to treat people with specialized needs, such as persons with co-occurring psychiatric conditions, adolescents, older adults, and persons with HIV and liver disease. Each of these conditions imposes specific requirements that must be addressed through careful patient selection, monitoring, and adjunctive services. There is little data on efficacy and safety to support the use of buprenorphine monotherapy to treat breast-feeding women. However, the benefits of breast-feeding are multiple, including a natural strengthening of the maternal-child bond, which is of particular importance for this patient population. The decision to place breast-feeding women on buprenorphine requires a balance of the risks and benefits and a discussion between provider and the individual patient, as a lack of data does not necessarily imply a risk of harm to the neonate. Populations not discussed by this panel remain of importance in treating opioid dependence, including

homeless populations, homosexual and bisexual populations, as well as patients with other medical comorbidities. Though beyond the scope of this review, these populations should be considered individually regarding treatment plans for opioid addiction.

CONCLUSIONS

Based on a review of the available evidence by a consensus panel with considerable clinical expertise and experience in the use of buprenorphine, the following recommendations are offered:

1. Medication-assisted therapies such as buprenorphine have been shown to be more effective than any other type of treatment for opioid dependence, particularly when used in concert with psychosocial interventions, such as counseling and other psychosocial therapies.
2. As in the treatment of most chronic diseases, pharmacotherapy of opioid dependence should be expected to take place over an extended period of time to achieve continued effective management of the underlying disorder. Most successful patients receive maintenance medication for years, whereas only a minority successfully taper off medication.
3. Therapeutic outcomes for patients who self-select office-based treatment with buprenorphine are essentially comparable to those seen in patients treated with methadone programs. While data thus far demonstrate that buprenorphine is almost as effective as methadone at promoting treatment retention, reducing illicit use of heroin and prescription opioids, reducing risky behaviors that transmit HIV and hepatitis, and is superior to methadone in terms of safety, further clinical data are still needed.
4. There are few absolute contraindications to the use of buprenorphine. However, the experience and skill levels of treating physicians can vary considerably, as can access to the resources needed to treat comorbid medical or psychiatric conditions—all of which may make the use of buprenorphine more complex.
5. Some patients who could benefit from treatment with buprenorphine may face challenges to successful treatment through office-based care, either because they require the structure afforded by a methadone program or because they lack access to office-based treatment or such care is not covered. It is important to conduct a targeted assessment of every patient to confirm that the provider has resources available to meet the patient's needs.
6. Patients should be assessed for a broad array of biopsychosocial needs in addition to opioid use and addiction, and should be treated, referred for help, or both in meeting all their care needs, including medical care, psychiatric care, and social assistance. In addition to the benefit they obtain directly from medication, patients should be encouraged to develop relapse prevention skills and to make active changes in their life circumstances to reduce relapse risk.
7. Most patients are likely to stabilize on 8 to 24 mg of buprenorphine per day, although some may need doses of up to 32 mg/day. In the absence of specific contraindications, the buprenorphine/naloxone combination is preferred to the monoproduct.
8. Although drug interactions with buprenorphine do occur, they are not always clinically relevant in a particular patient and do not necessarily prohibit the concomitant administration of buprenorphine with other drugs (although adjustment of the buprenorphine dose may be necessary). In any case, patients should be informed of the potential for drug interactions.
9. Physicians who wish to use buprenorphine should seek a level of comfort with this treatment approach. This encompasses knowledge of applicable practice standards and guidelines, familiarity with the evidence supporting recommended treatment strategies, protocols for treatment or referral of patients with complicating conditions (eg, severe depression, pain, or pregnancy), and an understanding of applicable federal, state, and local laws and regulations.
10. Physicians who treat opioid-dependent patients with buprenorphine should engage in continued medical education and other professional activities to keep current with the evolving knowledge base regarding optimal use of medication-assisted therapies in general with a particular focus on buprenorphine.

ACKNOWLEDGMENTS

The consensus panel and national peer review panel acknowledge with gratitude the support and encouragement of the Officers and board of directors of the American Society of Addiction Medicine, and of ASAM's Executive Vice President, Eileen McGrath, JD. Deep appreciation also is due consensus panel chair Mark L. Kraus, MD, FASAM, for his unwavering leadership, and to ASAM staff members Tracy Gartenmann, Angela Warner, and Darlene Williams for the exceptional skill with which they administered the project.

Consensus Panel. *The consensus panel developed the findings, conclusions, and recommendations found in this report. Members of the panel are:*

Daniel P. Alford, MD, MPH, FACP, Associate Professor of Medicine, Boston University School of Medicine, and Medical Director, Office-Based Opioid Treatment and MASBIRT Programs, Boston Medical Center, Boston, Massachusetts

Margaret M. Kotz, DO, Professor of Psychiatry, Case Western Reserve University School of Medicine, and Director of Addiction Recovery Services, University Hospitals of Cleveland, Cleveland, Ohio

Mark L. Kraus, MD, FASAM, Assistant Clinical Professor of Medicine, Yale University School of Medicine, New Haven, Connecticut; Primary Care General Internal Medicine/Addiction Medicine, Westside Medical Group, PC, Waterbury, Connecticut, and Chief Medical Officer, Connecticut Counseling Centers, Danbury, Connecticut

Petros Levounis, MD, MA, FASAM, Director, The Addiction Institute of New York, and Chief, Division of Addiction

Psychiatry, St Luke's & Roosevelt Hospitals, and Associate Clinical Professor of Psychiatry, Columbia University College of Physicians & Surgeons, New York, New York

Todd W. Mandell, MD, Chief Medical Officer, Community Substance Abuse Centers, Inc, Westfield, Massachusetts

Marjorie Meyer, MD, Director, Opiate Treatment Program During Pregnancy and Associate Professor, Ob/Gyn, University of Vermont Medical Center, Burlington, Vermont

Edwin A. Salsitz, MD, FASAM, Medical Director, Office-Based Opioid Therapy, Department of Medicine, Beth Israel Medical Center, New York, New York

Norman Wetterau, MD, FAAFP, FASAM, Private Practice of Family Medicine, Tri-County Family Medicine, Dansville, New York

Stephen A. Wyatt, DO, Medical Director, Dual Diagnosis Program, Middlesex Hospital, Middletown, Connecticut

National Peer Review Panel. Members of the national peer review panel reviewed multiple drafts of this article and supporting documents and made numerous suggestions to improve its clarity, scientific accuracy, and clinical utility. The consensus panel is indebted to the peer reviewers for their many contributions.

Peter Banys, MD, MSc, Health Sciences Clinical Professor of Psychiatry, University of California at San Francisco, and Director, Substance Abuse Program & Substance Abuse Physician Fellowship, VA Medical Center, San Francisco, California

Louis E. Baxter, Sr, MD, FASAM, Executive Medical Director, Professional Assistance Program of New Jersey, and Clinical Professor of Medicine, University of Medicine and Dentistry of New Jersey, and President-Elect, American Society of Addiction Medicine, Princeton, New Jersey

Anthony Dekker, DO, FAOAAM, FASAM, Chief Clinical Consultant for Addiction Medicine, Indian Health Service, and Phoenix Indian Health Center, Phoenix, Arizona

John Femino, MD, President and Medical Director, Meadows Edge Recovery Center, North Kingstown, Rhode Island

David A. Fiellin, MD, Associate Professor of Medicine and Investigative Medicine, Yale University School of Medicine, New Haven, Connecticut

Michael Fleming, MD, MPH, Professor of Family Medicine, University of Wisconsin-Madison Medical School, Madison, Wisconsin

Peter D. Friedmann, MD, MPH, Division of General Medicine, Rhode Island Hospital, Providence, Rhode Island

Kathleen Gargano, RN, National Association of Advocates for Buprenorphine Treatment, Farmington, Connecticut

Angel A. González, MD, Private Practice of Medicine, San Juan, Puerto Rico

Eric Goplerud, PhD, Center for Integrated Behavioral Health Policy, George Washington University Medical Center, Washington, DC

J. Harry Isaacson, MD, Associate Professor of Medicine and Director of Clinical Education, Department of General Internal Medicine, Cleveland Clinic Lerner College of Medicine, Cleveland, Ohio

Herbert D. Kleber, MD, Director, Division on Substance Abuse, Columbia University/NYSPI, New York, New York

Judith Martin, MD, Medical Director, BAART Turk Street Clinic, San Francisco, California

Karen Miotto, MD, Director of the UCLA Addiction Medicine Service and Clinical Professor, David Geffen School of Medicine at UCLA, Semel Institute of Neuroscience, Los Angeles, California

A. Kenison Roy III, MD, FASAM, President and Medical Director, Addiction Recovery Resources, Metairie, Louisiana

Samuel M. Silverman, MD, Private Practice of Medicine, West Hartford, Connecticut

Richard G. Soper, MD, JD, MS, Center for Integrated Healthcare, Nashville, Tennessee

Lynn E. Sullivan, MD, Assistant Professor of Medicine, Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut

Trusandra E. Taylor, MD, MPH, Medical Director, JEVS Human Services, Philadelphia, Philadelphia

Berton Toews, MD, Medical Director, Wyoming Recovery LLC, Casper, Wyoming

John J. Verdon, Jr, MD, FASAM, Private Practice of Medicine, Shrewsbury, New Jersey

George E. Woody, MD, Professor, Department of Psychiatry, University of Pennsylvania, and Treatment Research Institute, Philadelphia, Pennsylvania

Martha J. Wunsch, MD, Center on Drug & Alcohol Research, Department of BehavioralScience, University of Kentucky School of Medicine, Lexington, Kentucky

Douglas M. Ziedonis, MD, MPH, Professor and Chairman, Department of Psychiatry, University of Massachusetts Medical School/UMass Memorial Health Care, Worcester, Massachusetts

REFERENCES

Alford DP, Compton P, Samet JH. Acute pain management for patients receiving maintenance methadone or buprenorphine therapy [Review]. *Ann Intern Med* 2006;144(2):127-134. Erratum in *Ann Intern Med* 2006;144(6):460.

Amato L, Minozzi S, Davoli M, Vecchi S, Ferri MM, Mayet S. Psychosocial and pharmacological treatments versus pharmacological treatments for opioid detoxification. *Cochrane Database Syst Rev* 2008a;(4):CD005031. (Update of *Cochrane Database Syst Rev* 2004 18;(4):CD005031).

Amato L, Minozzi S, Davoli M, Vecchi S, Ferri MM, Mayet S. Psychosocial combined with agonist maintenance treatments versus agonist maintenance treatments alone for treatment of opioid dependence. *Cochrane Database*

Syst Rev 2008b; (4):CD004147. (Update of *Cochrane Database Syst Rev* 2004;(4):CD005031).
Amato L, Davoli M, Perucci CA, Ferri M, Faggiano F, Mattick RP. An overview of systematic reviews of the effectiveness of opiate maintenance therapies: available evidence to inform clinical practice and research. *J Subst Abuse Treat* 2005;28:321–329.
Anderson IB, Kearney TE. Use of methadone. *West J Med* 2000;172(1):43–46.
Backmund M, Meyer K, Von Zielonka M, Eichenlaub D. Treatment of hepatitis C infection in injection drug users. *Hepatology* 2001;34(1):188–193.
Batki SL, for the CSAT Consensus Panel. Medication-Assisted Treatment for Opioid Addiction in Opioid Treatment Programs (Treatment Improvement Protocol [TIP] Series, Number 43). Rockville: Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration (DHHS Publication No. [SMA] 06-4214), 2005.
Baxter LE for the ASAM Methadone Action Group. Clinical Guidance on Methadone Induction and Stabilization. Chevy Chase: American Society of Addiction Medicine, Inc (in field review 2009).
Becker WC, Sullivan LE, Tetrault JM, Desai RA, Fiellin DA. Non-medical use, abuse and dependence on prescription opioids among U.S. adults: psychiatric, medical and substance use correlates. *Drug Alcohol Depend* 2008;94(1–3):38–47.
Berson A, Gervais A, Cazals D, et al. Hepatitis after intravenous buprenorphine misuse in heroin addicts. *J Hepatol* 2001;34:346–350.
Bio LL, Siu A, Poon CY. Update on the pharmacologic management of neonatal abstinence syndrome. *J Perinatol* August 25, 2011 [epub ahead of print].
Blow F. Older adults. In: Graham AW, Schultz TK, Mayo-Smith MF, et al., eds. Principles of Addiction Medicine, Third Edition. Chevy Chase: American Society of Addiction Medicine, 2003.
Briggs GC, Freeman RK, Yaffe SJ. Drugs in Pregnancy and Lactation, Eight Edition. Philadelphia: Lippincott Williams & Wilkins, 2008: 212–213.
Bruce RD, Altice FL. Three case reports of a clinical pharmacokinetic interaction between buprenorphine and atazanavir plus ritonavir. *AIDS* 2006;20(5):783–784.
Caplehorn J, Deeks JJ. A critical appraisal of the Australian comparative trial of methadone and buprenorphine maintenance [Review]. *Drug Alcohol Rev* 2006;25(2):157–160.
Carrieri MP, Vlahov D, Dellamonica P, et al. Use of buprenorphine in HIV infected injection drug users: negligible impact on virological response to HAART. The Manif-2000 Study Group. *Drug Alcohol Depend* 2000;60: 51–54.
Cazorla C, Grenier de Cardenal D, Schumacher H, et al. Infectious complications and misuse of high-dose buprenorphine. *Presse Med* 2005;34(10): 719–724.
Center for Substance Abuse Treatment. Substance Abuse: Administrative Issues in Outpatient Treatment (Treatment Improvement Protocol [TIP] 46). Rockville: Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration, 2006a.
Center for Substance Abuse Treatment. The SAMHSA Evaluation of the Impact of the DATA Waiver Program: Summary Report. Rockville: Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration, 30, 2006b.
Collins GB, McAllister MS. Buprenorphine maintenance: a new treatment for opioid dependence. *Cleve Clin J Med* 2007;74(7):514–520.
Comer SD, Sullivan MA, Walker EA. Comparison of intravenous buprenorphine and methadone self-administration by recently detoxified heroin-dependent individuals. *J Pharmacol Exp Ther* 2005a;315(3): 1320–1330.
Comer SD, Walker EA, Collins ED. Buprenorphine/naloxone reduces the reinforcing and subjective effects of heroin in heroin-dependent volunteers. *Psychopharmacology* (Berl) 2005b;181(4):664–675.
Connors GJ, Longabaugh R, Miller WR. Looking forward and back to relapse: implications for research and practice. *Addiction* 1996;91 (Suppl):S191–S196.
Daley DC, Salloum IM, Zuckoff A, Kirisci L, Thase ME. Increasing treatment adherence among outpatients with depression and cocaine dependence: results of a pilot study. *Am J Psychiatry* 1998;155(11):1611–1613.
Dashe JS, Sheffield JS, Wendel GD Jr. Improving the management of opioid-dependent pregnancies. *Am J Obstet Gynecol* 2004;190(6):1806; author reply 1806–1807.
Dausey DJ, Desai RA. Psychiatric comorbidity and the prevalence of HIV infection in a sample of patients in treatment for substance abuse. *J Nerv Ment Dis* 2003;191(1):10–17.
Dimeff LA, Marlatt GA. Relapse prevention. In: Hester R, Miller W, eds. Handbook of Alcoholism Treatment Approaches, 2nd Edition. Boston: Allyn & Bacon, 1995:176–194.
Edge WG, Cooper GM, Morgan M. Analgesic effects of sublingual buprenorphine. *Anaesthesia* 1979;34(5):463–467.
Fiellin DA. The first three years of buprenorphine in the United States: experience to date and future directions. *J Addict Med* 2007;1(2):62–67.
Fiellin DA. Treatment of adolescent opioid dependence: no quick fix. *JAMA* 2008; 300(17):2057–2059.
Fiellin DA, Friedland GH, Gourevitch MN. Opioid dependence: rationale for and efficacy of existing and new treatments [Review]. *Clin Infect Dis* 2006;43(Suppl 4):S173–S177.
Fiellin DA, Moore BA, Sullivan LE, et al. Long-term treatment with buprenorphine/naloxone in primary care: results at 2-5 years. *Am J Addict* 2008;17(2):116–120.
Fishman MJ, Mee-Lee D, Shulman GD, et al. Supplement to the ASAM Patient Placement Criteria on the Management of Alcohol Use Disorders. Baltimore: Lippincott Williams & Wilkins, 2010.
Fleming MF, Davis J, Passik SD. Reported lifetime drug-taking behaviors are predictive of current substance use and mental health problems in primary care patients. *Pain Med* 2008;9(8):1098–1106.
Gordon AJ, Krumm M, for the Buprenorphine Initiative in the VA (BIV). Buprenorphine Resource Guide, Version 8. Washington: Department of Veterans Affairs, 2008.
Gowing L, Ali R, White J. Buprenorphine for the management of opioid withdrawal [Review]. *Cochrane Database Syst Rev* 2006;(2):CD002025.
Gowing L, Ali R, White J. Buprenorphine for the management of opioid withdrawal [Review]. *Cochrane Database Syst Rev* 2009;(3):CD002025-pub4.
Gunderson EW, Fiellin DA. Office-based maintenance treatment of opioid dependence: how does it compare with traditional approaches? *CNS Drugs* 2008;22(2):99–111.
Han B, Gfroerer JC, Colliver JD, Penne MA. Substance use disorder among older adults in the United States in 2020. *Addiction* 2009;104(1):88–96.
Health Resources and Services Administration. An Evaluation of Innovative Methods for Integrating Buprenorphine Opioid Abuse Treatment in HIV Primary Care Settings: (1) Demonstration Models and (2) Evaluation and Support Center (HRSA-04-078). Rockville: HIV/AIDS Bureau, Demonstration Project Development and Evaluation Branch, Special Projects of National Significance, 2004. Available at: http://hab.hrsa.gov/abouthab/special/buprenorphine.html. Accessed October 23, 2006.
Helm S, Trescot AM, Colson J, Sehgal N, Silverman S. Opioid antagonists, partial agonists, and agonists/antagonists: the role of office-based detoxification [Review]. *Pain Physician* 2008;11(2):225–235.
Johnson RE, Jones HE, Fischer G. Use of buprenorphine in pregnancy: patient management and effects on the neonate. *Drug Alcohol Depend* 2003;70(Suppl 2):S87–S101.
Johnson RE, Strain EC, Amass L. Buprenorphine: how to use it right. *Drug Alcohol Depend* 2003;70(Suppl):S59–S77.
Jones HE, Kaltenbach K, Heil SH, et al. Neonatal abstinence syndrome after Methadone or Buprenorphine exposure. *N Engl J Med* 2010;363:2320–2331.
Jones HE, Jasinski DR, O'Grady KE, et al. Buprenorphine versus methadone in the treatment of pregnant opioid-dependent patients: effects on the neonatal abstinence syndrome. *Drug Alcohol Dependence* 2005;79(1):1–10.
Jones HE, Martin PR, Heil SH, et al. Treatment of opioid-dependent pregnant women: clinical and research issues. *J Subst Abus Treat* 2008;35(3):245–259.
Joseph H, Stancliff S, Lagrod J. Methadone maintenance treatment (MMT): a review of historical and clinical issues [Review]. *Mt Sinai J Med* 2000;67(5–6):347–364.
Kakka J, Hellig M, Sarman I. Buprenorphine and methadone treatment of opiate dependence during pregnancy: comparison of fetal growth and neonatal outcomes in two consecutive case series. *Drug Alcohol Depend* 2008;96 (1-2):69–78.
Kalivas PW, Volkow ND. The neural basis of addiction: a pathology of motivation and choice. *Am J Psychiatry* 2005;162:1403–1413.
Kessler RC, McGonagle KA, Zhao S, et al. Lifetime and 12-month prevalence of DSM-IIIR psychiatric disorders in the United States. *Arch Gen Psychiatry* 1994;51:8–19.

Khalsa JH, Treisman G, McCance-Katz EF, Tedaldi E. Medical consequences of drug abuse and co-occurring infections: research at the National Institute on Drug Abuse. *Subst Abus* 2008;29(3):5–16.

Kleber HD. Pharmacologic treatments for opioid dependence: detoxification and maintenance options [Review]. *Dialogues Clin Neurosci* 2007;9(4):455–470.

Kraft WK, Dysart K, Greenspan JS, Gibson E, Kaltenbach K, Ehrlich ME. Revised dose schema of sublingual buprenorphine in the treatment of the neonatal opioid abstinence syndrome. *Addiction* 2011;106:574–580.

Lavie E, Fatseas M, Denis C, Auriacombe M. Benzodiazepine use among opiate-dependent subjects in buprenorphine maintenance treatment: correlates of use, abuse and dependence. *Drug Alcohol Depend* 2009;99(1–3):338–344.

Leavitt SB, Shinderman M, Maxwell S, Eap CB, Paris P. When "enough" is not enough: new perspectives on optimal methadone maintenance dose [Review]. *Mt Sinai J Med* 2000;67(5–6):404–411.

Lejeune C, Simmat-Durand L, Gourarier L, Aubisson S; Groupe d'Etudes Grossesse et Addictions (GEGA). Prospective multicenter observational study of 260 infants born to 259 opiate-dependent mothers on methadone or high-dose buprenorphine substitution. *Drug Alcohol Depend* 2006;82(3):250–257.

Levy S, Vaughan BL, Angulo M, Knight JR. Buprenorphine replacement therapy for adolescents with opioid dependence: early experience from a children's hospital-based outpatient treatment program. *J Adolesc Health* 2007;40(5):477–488.

Marlatt G, Gordon JR, eds. Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors. New York: Guilford Press, 1985.

Marsch LA, Bickel WK, Badger GJ, et al. Comparison of pharmacological treatments for opioid-dependent adolescents. *Arch Gen Psychiatry* 2005;62:1157–1164.

Mattick RP, Kimber J, Breen C, Davoli M. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. *Cochrane Database Syst Rev* 2008;(2):CD002207 [Update of *Cochrane Database Syst Rev* 2004;(3): CD002207].

McCance-Katz EF. Treatment of opioid dependence and coinfection with HIV and hepatitis C virus in opioid-dependent patients: The importance of drug interactions between opioids and antiretroviral agents. *Clin Infect Dis* 2005; 41(Suppl 1):S89–S95.

McCance-Katz EF, Moody DE, Morse GD, et al. Interactions between buprenorphine and antiretrovirals. I. The nonnucleoside reverse-transcriptase inhibitors efavirenz and delavirdine. *Clin Infect Dis* 2006a;43(Suppl 4):S224–S234.

McCance-Katz EF, Moody DE, Smith PF, et al. Interactions between buprenorphine and antiretrovirals. II. The protease inhibitors nelfinavir, lopinavir/ritonavir, and ritonavir. *Clin Infect Dis* 2006b;43(Suppl 4):S235–S246.

McNicholas LM for the CSAT Consensus Panel. Clinical Guidelines for the Use of Buprenorphine in the Treatment of Opioid Addiction (Treatment Improvement Protocol [TIP] Series, Number 40). Rockville: Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration (DHHS Publication No [SMA] 04-3939), 2004.

McNicholas L. Clinical guidelines for the use of buprenorphine in the treatment of opioid addiction. *A Tool for Buprenorphine Care* 2008;1(12):12.

Moa G, Zetterström H. Sublingual buprenorphine as postoperative analgesic: a double-blind comparison with pethidine. *Acta Anaesthesiol Scand* 1990;34(1):68–71.

Moatti JP, Carrieri MP, Spire B, Gastaut JA, Cassuto JP, Moreau J. Adherence to HAART in French HIV-infected injecting drug users: the contribution of buprenorphine drug maintenance treatment. The Manif 2000 Study Group. *AIDS* 2000;4:151–155.

National Institute on Drug Abuse. *Principles of Drug Addiction Treatment: A Research-Based Guide* (NIH Publication No. 00-4180). Rockville: National Institute on Drug Abuse, National Institutes of Health, 2000.

Paulozzi LJ, Budnitz DS, Xi Y. Increasing deaths from opioid analgesics in the United States. *Pharmacoepidemiol Drug Saf* 2006;15(9):618–627.

Reed LJ, Glasper A, de Wet CJ, Bearn J, Gossop M. Comparison of buprenorphine and methadone in the treatment of opiate withdrawal: possible advantages of buprenorphine for the treatment of opiate-benzodiazepine codependent patients? *J Clin Psychopharmacol* 2007;27(2):188–192.

Robins LN, Locke BZ, Regier DA. An overview of psychiatric disorders in America. In: Robins LN, Regier DA, eds. Psychiatric Disorders in America: The Epidemiologic Catchment Area Study. New York: Free Press, 1991: 328–366.

Room R. The Co-Occurrence of Mental Disorders and Addictions: Evidence on Epidemiology, Utilization, and Treatment Outcomes (ARF Research Document Series, No. 141). Ontario: Addiction Research Foundation, Addiction and Mental Health Services Corporation, 1998.

Sacks S, Ries RR for the CSAT Consensus Panel. Substance Abuse Treatment for Persons with Co-Occurring Disorders. (Treatment Improvement Protocol [TIP] Series, Number 42). Rockville: Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration (DHHS Publication No. [SMA] 05-3922), 2005.

Sanchez ES, Bigbee JW, Fobbs W, Robinson SE, Sato-Bigbee C. Opioid addiction and pregnancy: perinatal exposure to buprenorphine affects myelination in the developing brain. *Glia*. 2008;56(9):1017–1027.

Schifano F, Corkery J, Gilvarry E, Deluca P, Oyefeso A, Ghodse AH. Buprenorphine mortality, seizures and prescription data in the UK, 1980–2002. *Hum Psychopharmacol* 2005;20(5):343–348.

Schindler SD, Eder H, Ortner R, Rohrmeister K, Langer M, Fischer G. Neonatal outcome following buprenorphine maintenance during conception and throughout pregnancy. *Addiction* 2003;98:103–110.

Soeffing JM, Martin LD, Fingerhood MI, Jasinski DR, Rastegar DA. Buprenorphine maintenance treatment in a primary care setting: outcomes at 1 year. *J Subst Abuse Treat* 2009;37(4):426–430.

Spiller H, Lorenz DJ, Bailey EJ, Dart RC. Epidemiological trends in abuse and misuse of prescription opioids. *J Addict Dis* 2009;28(2):130–136.

Sporer KA. Buprenorphine: a primer for emergency physicians. *Ann Emerg Med* 2004;43(5):580–584.

Sproule B, Brands B, Li S, Catz-Biro L. Changing patterns in opioid addiction: characterizing users of oxycodone and other opioids. *Can Fam Physician* 2009;55(1):68–69, 69.e1–e5.

Stephen DK for the CSAM Committee on Treatment of Opioid Dependence. Draft Guidelines for Physicians Working in California Opioid Treatment Programs. San Francisco: California Society of Addiction Medicine, 2008.

Stotts AL, Dodrill CL, Kosten TR. Opioid dependence treatment: options in pharmacotherapy. *Expert Opin Pharmacother* 2009;10(11):1727–1740.

Tetrault JM, Desai RA, Becker WC, Fiellin DA, Concato J, Sullivan LE. Gender and non-medical use of prescription opioids: results from a national US survey. *Addiction* 2008;103(2):258–268.

Tims F, Leukefeld C, eds. Relapse and Recovery in Drug Abuse (NIDA Research Monograph 72). Rockville: National Institute on Drug Abuse, National Institutes of Health, 1987.

Umbricht A, Huestis M, Cone EJ, Preston KL. Effects of high-dose intravenous buprenorphine in experienced opioid abusers. *J Clin Psychopharmacol* 2004;24(5):479–487.

Unger A, Jagsch R, Jones H, et al. Randomized controlled trials in pregnancy: scientific and ethical aspects. Exposure to different opioid medications during pregnancy in an intra-individual comparison. *Addiction* 2011;106:1355–1362.

Vavrinková B, Binder T. The effect of substitution therapy on the birth weight of the newborn, its postpartum adaptation, trophic and course of the neonatal abstinence syndrome. *Ceska Gynekol* 2007;72(4):247–253.

Wedam EF, Bigelow GA, Johnson RE, et al. QT-interval effects of methadone, levomethadyl, and buprenorphine in a randomized trial. *Arch Intern Med* 2007;167:2469–2475.

White W. Addiction recovery: its definition and conceptual boundaries. *J Subst Abuse Treat* 2007;33:229–241.

Woody GE, Poole SA, Subramaniam G, et al. Extended vs short-term buprenorphine-naloxone for treatment of opioid-addicted youth: a randomized trial. *JAMA* 2008;300(17):2003–2011.

World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. Geneva: World Health Organization, 2009.

Ziedonis D, Steinberg ML, Smelson D, et al. Co-occurring addictive and psychotic disorders. In: Graham AW, Schultz TK, Mayo-Smith MF, et al., eds. Principles of Addiction Medicine, Third Edition. Chevy Chase: American Society of Addiction Medicine, Inc, 2003.

BMJ Quality & Safety Online First, published on 22 September 2011 as 10.1136/bmjqs-2011-000084

Factors associated with disclosure of medical errors by housestaff

Andrea C Kronman,[1,2] Michael Paasche-Orlow,[1] Jay D Orlander[1,3]

► An additional appendix is published online only. To view this file please visit the journal online (http://qualitysafety.bmj.com).

[1]Section of General Internal Medicine, Evans Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA
[2]Women's Health Interdisciplinary Research Center, Boston University School of Medicine, Boston, Massachusetts, USA
[3]Medical Service, VA Boston Healthcare System, Boston, Massachusetts, USA

Correspondence to
Dr Andrea C Kronman, Boston University Medical Center, 801 Massachusetts Ave, Suite 470, Boston, MA 02118, USA;
andrea.kronman@bmc.org

Accepted 16 August 2011

ABSTRACT

Purpose: Attributes of the organisational culture of residency training programmes may impact patient safety. Training environments are complex, composed of clinical teams, residency programmes, and clinical units. We examined the relationship between residents' perceptions of their training environment and disclosure of or apology for their worst error.

Method: Anonymous, self-administered surveys were distributed to Medicine and Surgery residents at Boston Medical Center in 2005. Surveys asked residents to describe their worst medical error, and to answer selected questions from validated surveys measuring elements of working environments that promote learning from error. Subscales measured the microenvironments of the clinical team, residency programme, and clinical unit. Univariate and bivariate statistical analyses examined relationships between trainee characteristics, their perceived learning environment(s), and their responses to the error.

Results: Out of 109 surveys distributed to residents, 99 surveys were returned (91% overall response rate), two incomplete surveys were excluded, leaving 97: 61% internal medicine, 39% surgery, 59% male residents. While 31% reported apologising for the situation associated with the error, only 17% reported disclosing the error to patients and/or family. More male residents disclosed the error than female residents ($p=0.04$). Surgery residents scored higher on the subscales of safety culture pertaining to the residency programme ($p=0.02$) and managerial commitment to safety ($p=0.05$). Our Medical Culture Summary score was positively associated with disclosure ($p=0.04$) and apology ($p=0.05$).

Conclusion: Factors in the learning environments of residents are associated with responses to medical errors. Organisational safety culture can be measured, and used to evaluate environmental attributes of clinical training that are associated with disclosure of, and apology for, medical error.

INTRODUCTION

Everyone makes mistakes. Over the past decade, the medical profession has started to apply a systems approach to patient safety, recognition that coordination of individual, team, and organisational forces are needed to promote patient safety. Analysis of the root causes of an error can prevent future errors by identifying and correcting problems.[1] However, in order to learn from mistakes and develop safer systems, errors must first be identified and reported.

Unfortunately, many errors are never reported. In one study, merely half of the house officers told their attending physicians about the most serious errors they committed.[2] Underreporting of adverse events is estimated to range from 50% to 96% annually.[1 3 4] Rather than dealing with mistakes constructively by reporting and learning from them, studies indicate that physicians typically respond to their mistakes defensively, blaming the system, other members of the healthcare team, or even the patient.[2 5–7] Possible explanations for underreporting medical errors include fear of litigation acting as a deterrent,[8–10] and the professional medical culture that limit an individual's willingness to discuss error.[5 11]

While elements of professional medical culture are hypothesised to lead to widespread underreporting of medical errors, few studies have elucidated and measured aspects of medical culture that are associated with a failure to disclose, particularly in the learning environments of clinical training programmes. In contrast to medical culture, non-medical industries such as aviation and nuclear safety have traditionally valued a professional 'culture of safety', which facilitates reporting of errors, so that individuals operating in groups within an organisation can learn how to prevent future errors.[12] Medical educators have recently attempted to incorporate system-based thinking into their curriculum, in order to incorporate aspects of a safety culture that, along with enquiry and trust, were previously lacking in

residency settings.[2 7 12] Although there is recognition of the need to create a learning culture of safety for residents,[13] measuring educational culture has proved to be a challenge. In addition to the complexities of quantitatively measuring an abstract concept of 'organisational culture',[14] residents train in multiple environments which are dynamic and divergent: their individual clinical teams (which often rotate), the academic residency training programme, and the clinical institution(s) (usually a hospital) each contributing to a trainee's overall sense of culture.

This study endeavours to measure a trainee's perception of their training environment as it relates to safety, and their response to committing an error. By collecting data on both these elements, we explore the relationship between the culture of the training environment and individual behaviour. Specifically, we hypothesised that house officers would be more likely to disclose and apologise for an error if they rate their clinical team as having an environment in which they can report errors without fear of punishment or rejection; rate their training programme as having positive attitudes about reporting and coping with errors in the workplace; and rate the hospital as having a high level of commitment to patient safety. Secondary aims of our study were to evaluate the association between individual characteristics of clinical trainees and, first, disclosure of a medical error, and second, apology for the error.

METHODS

Study design and survey administration

Anonymous, self-administered surveys were distributed to medicine and surgery residents at Boston Medical Center during educational conferences and department meetings in 2005. The distribution and retrieval procedures of the surveys ensured privacy and anonymity of the residents. The researchers (who were also attending physicians) were blinded as to which residents completed the survey, which ensured no conflict of interest for the researchers if asked to evaluate residents. The residents were assured that their privacy and anonymity would be protected, and that the researchers would remain blinded to their participation status. Specifically, the researchers approached a group of housestaff during an educational conference or meeting, explained the purpose of the survey, and then left the room and building. Residents who choose to participate completed the surveys and returned them in sealed envelopes to a box in the room. Those who chose not to participate returned blank surveys in sealed envelopes. At the end of each conference, a research assistant returned to the room to collect the box with the sealed envelopes. All participants received a $10 honorarium, whether or not they completed the survey. The survey, database, and protocol were de-identified. To further protect participants in the event of an accidental breach of anonymity, a certificate of confidentiality was obtained from the National Institutes of Health. The project was carefully reviewed and approved by the institutional review board.

Survey content

Our survey focused on three levels of environment that had the most face validity of microculture constructs within a resident's learning environment. Questions were selected from three validated surveys of organisational culture, adapted for this study to focus on the organisational environment of housestaff. Since the full survey instruments were deemed too high a respondent burden, the authors carefully considered and then selected items from each survey most relevant to the study. The microenvironment of the immediate clinical team was examined with five of seven questions from the Team Psychological Safety Survey, which assesses the belief that well intentioned actions will not lead to punishment or rejection by the team (see online appendix).[15] The macroenvironment of the residency programme was assessed with 10 of the 37 items from the Error Orientation Questionnaire; the items selected assess attitudes to errors and approaches to coping with errors in the workplace (see online appendix).[16] Perception of hospital management's commitment to patient safety on the clinical unit was assessed with four of 19 items from the Patient Safety Survey (see online appendix).[17] Responses were coded with six-point Likert scales, and summed to derive a total score for each survey. For all three scales (Team Psychological Safety, Residency Programme Error Orientation, and Managerial Safety Commitment), higher scores correlated with more positive aspects of culture. In relation to error, participants were asked to recall the circumstances of and share details regarding their most significant medical or surgical error using open-ended text. In a multiple choice format, participants were specifically asked about the following: consequences for the patient; consequences for the resident; if and to whom they disclosed the error; if they had apologised for the error; and perceived causes for the error. Responses to each question were constructed from the results of a previous survey of residents regarding medical error, conducted by Wu *et* al.[2] In addition, residents were asked to characterise their own level of distress from the error using a 10-point Likert scale.

Analysis

Univariate analysis was used to describe demographics, residency type, reporting rates (to colleagues and friends), apology and disclosure rates (to patients),

emotional responses of residents, types of mistakes, and consequences to and responses of both the residents and patients. Errors were classified from the written responses into one of the following categories: procedural, medical management, laboratory test follow-up, delayed diagnosis, or other/not classifiable. We used χ^2 to evaluate differences between categorical variables and Wilcoxon rank sum methods for the three organisational culture scales. In order to compare the subscales with each other, the raw score was converted to a scaled score, by dividing each raw score by the maximum possible score of each subscale, and multiplying by 33.3. We then calculated an overall Medical Culture Summary score, by summing the three scaled subscores, that is, each subscale contributes one-third of the overall Medical Culture Summary score.

RESULTS

Surveys were distributed to 109 residents and 99 surveys were returned, making an overall response rate of 91%. Two residents' surveys were excluded because they reported no mistake, leaving a final population of 97 residents, 59 (61%) from internal medicine residents and 38 (39%) from surgical residents. There were 57 (59%) male residents, of which 33 (58%) were internal medicine residents and 24 (42%) were surgical residents. Two surgical residents did not report their gender, and were excluded from analyses which included gender. The most significant medical or surgical error that was the focus of residents' responses typically occurred in an inpatient setting and during the first year of training (table 1).

Table 1 Characteristics of residents and setting where error occurred

	Total N=97 N (%)
Gender	
Men	57 (59)
Women	38 (39)
Unknown	2 (2)
Programme	
Medicine	59 (61)
Men	33 (56)
Surgery	38 (39)
Men	24 (63)
Training year	
First year	64 (66)
After first year	32 (33)
Unknown	1 (1)
Setting	
Ward	63 (65)
Ambulatory clinic	3 (3)
Operating room	5 (5)
Intensive care unit	22 (23)
Emergency department	4 (4)

Seventy-five per cent of the residents were extremely distressed by their mistake. While 41 (42%) did not provide an adequate description of their error to be classified, 26 (27%) were classified as medication related, 12 (12%) as procedural, 11 (11%) due to delayed diagnosis, and 9 (9%) due to inadequate follow-up to a laboratory test.

Although 20 (21%) of the involved patients had no reported consequences resulting from the errors, common consequences included delayed treatment for 23 (24%), delayed diagnosis for 22 (23%), prolonged hospital stay for 17 (18%), medical complications for 13 (13%), and death for 13 (13%) patients. The errors resulting in patient death were largely errors involving anticoagulants, potassium balance (either not checking blood work or inadequate management of blood potassium level), or insulin. There were no consequences for 60 (62%) of the residents due to the error, but 30 (30%) reported some form of reprimand, 16 (16%) presented the case at a morbidity and mortality conference (which was reported as a consequence), 6 (6%) reported their work and family life was affected, and 1 (1%) was named in a law suit. The most common attributions for the error reported by residents included being too busy (32, 33%) and inexperience (31, 32%). Many residents also attributed their error to having inadequate knowledge, hesitating before acting, or being too tired (table 2). While 30 (31%) reported apologising for the situation associated with the error, only 17 (18%) reported disclosing the error to patients and/or their family. Five residents both disclosed (29% of those who disclosed) and apologised (17% of those who apologised table 3).

Correlates of disclosure and apology

The disclosure rate was higher among surgery residents (24%) than internal medicine residents (14%), but this difference was not statistically significant (p=0.2). Of the residents who disclosed their error, 32 (33%) reported that it was unsupervised. Three (3%) residents reported being told by their attending not to discuss the error with the patient. Female internal medicine residents were significantly less likely to disclose their worst medical error to patients or their families than their male counterparts (p=0.03). In contrast, more female surgery residents, 7 (58%), apologised for their error compared with male surgery residents 7 (29%), though the difference did not reach statistical significance (p=0.1). Of the 13 errors that resulted in a patient's death, only 3 (23%) of the residents disclosed the error to the patient and/or the patient's family, but 6 (46%) residents apologised to the patient's family. More residents who made errors in medication management (8 of 26, 31%) disclosed their error than those who made

Table 2 Characteristics of worst medical errors*

	Total N=97N (%)
Cause of error (self-report)	
Too busy	32 (33)
Inexperience	31 (32)
Inadequate knowledge	19 (20)
Other	15 (15)
Hesitated too long before acting	14 (14)
Too tired	13 (13)
Inadequate communication	10 (10)
Inadequate supervision	9 (9)
Given erroneous information	9 (9)
Felt overwhelmed	8 (8)
Did not obtain the appropriate follow-up test	4 (4)
Inadequate history	3 (3)
Could not recall knowledge	2 (2)
Type of error (classified by investigator)	
Other/could not be classified	41 (42)
Medical management	26 (27)
Procedural	12 (12)
Delayed diagnosis	11 (11)
Did not check lab test	9 (9)
Consequences of error to patient	
Delayed treatment	23 (24)
Delayed diagnosis	22 (23)
None	20 (21)
Prolonged hospital stay	17 (18)
Medical complications	13 (13)
Death	13 (13)
Temporary symptoms	9 (9)
Required additional medications	9 (9)
Other	8 (8)
Additional imaging tests	6 (6)
Additional blood test	5 (5)
Stressed relationship with medical providers	1 (1)
Consequences to resident	
None	60 (62)
Present at M&M conference	16 (16)
Reprimanded by another resident	16 (16)
Reprimanded by attending	12 (12)
Work life affected	6 (6)
Personal life affected	6 (6)
Other	3 (3)
Reprimanded by administrator	2 (2)
Named in lawsuit	1 (1)
Error disclosure/discussion occurred with:	
Fellow residents	66 (68)
Supervising resident	48 (49)
Supervising attending	43 (44)
Significant other	22 (23)
Patient and/or patient's family	17 (18)
Colleagues at a conference	16 (16)
Relative or friend	8 (8)
No one	6 (6)
Other	3 (3)
If disclosure occurred to patient and/or family, who supervised?	
No one	32 (33)
Attending	18 (19)
Senior resident	14 (14)
Risk management	3 (3)
Other	1 (1)

*Residents could select more than one option.
M&M, morbidity and mortality.

errors with procedures (3 of 12, 25%) or delayed diagnosis (2 of 11, 18%) (p=0.05).

Correlates of organisational culture measures

The Safety Culture Summary score was positively associated with disclosure of medical error to the patient and/or patient's family (p=0.04) and apology for the error (p=0.05). There was a trend of association between disclosure and higher scores on the subscales clinical Team Psychological Safety (p=0.07) and Residency Programme Error Orientation scales (p=0.07), but not for Managerial Safety Commitment (p=0.2). Report of apology to the patient and/or patient's family was not associated with the clinical Team Psychological Safety score (p=1.0) but was positively associated with scores on the Residency Programme Error Orientation (p=0.05) and Managerial Safety Commitment (p=0.01). There were no significant gender differences in scores for each of the subscales as well as the summary measure of safety culture. Surgery residents had higher scores on the Residency Programme Error Orientation (p=0.02) and Managerial Safety Commitment scales (p=0.05) compared with medicine residents, but there was no significant difference between programmes in the Safety Culture Summary score (table 4).

CONCLUSIONS

Only 17% of the residents we surveyed reported disclosing their most significant error to their patient and/or patient's family, and only 31% of the residents reported apologising for their most significant error. Our results suggest that factors in the learning environments of the clinical team and residency programme are associated with error disclosure and apology among residents. Individual factors, such as gender and type of error, also appear to be associated with error disclosure

Table 3 Individual factors associated with disclosure and apology, by gender* and programme

	Yes (%)	p Value†
Disclosure to patient and/or patient's family?		
Total (N=97)	17 (18)	0.04
Men (N=57)	14 (25)	
Women (N=38)	3 (8)	
Programme		0.16
Medicine (N=59)	8 (14)	0.05
Men (N=33)	7 (21)	
Women (N=26)	1 (4)	
Surgery (N=38)	9 (24)	0.40
Men (N=24)	7 (29)	
Women (N=12)	2 (17)	
Apologise to patient and/or patient's family?		
Total (N=97)	30 (31)	0.6
Men (N=57)	17 (30)	
Women (N=38)	13 (34)	
Programme		0.2
Medicine (N=59)	16 (27)	0.5
Men (N=33)	10 (30)	
Women (N=26)	6 (23)	
Surgery (N=38)	14 (39)	0.09
Men (N=24)	7 (29)	
Women (N=14)	7 (58)	

*N=95; two respondents who did not enter gender were excluded from the analysis.
†p Values calculated using the χ^2 test.

and apology, and more residents apologised for the error than disclosed it.

Our findings of discordance between apology and disclosure of medical error are consistent with previous studies exhibiting residents may be more willing to apologise for a bad outcome than to reveal that they played a role in causing the bad outcome, resulting in a 'partial disclosure'.[18 19] Collectively, these findings imply that factors that facilitate apologising for an error may differ from influences that facilitate disclosing an error. These findings are reflected in State laws that distinguish different components of conversations with patients about unanticipated outcomes: 'expression of sympathy' (apology), 'explanation' (disclosure), and 'admission of fault', which does not cleanly translate into either category.[20] Additional explanations for the discordance may include the social context in which the error occurred. For example, apologising for a systemic error that occurred would likely be easier than disclosing personal responsibility for an error, which could have greater legal and professional consequences.[19]

The relationships of gender to our outcome measures are complex. More male residents disclosed error (driven mostly by male internal medicine residents) while more women apologised (driven mostly by female surgical residents). With our small sample size, definitive conclusions about the interactions among gender, specialty and disclosure are difficult to ascertain from our data. Previous studies[2 9] have also demonstrated that individual attributes, such as gender and emotional response to the error, influence the reporting rate of the error. However, in contrast to our results, women in a previous study were *more* likely to discuss their errors with their patients and make constructive changes in their practice.[2] Although the female residents in our study were less likely to disclose their error, the female surgical residents were more likely to apologise, consistent with past reports of greater empathy among female physicians.[21] Our results suggest that there are barriers to disclosure in the learning environments of clinical trainees that affect men and women differently. Further research will be needed to elucidate which barriers to disclosure and apology affect genders differentially. For example, possible barriers to disclosure may be attitudinal— women may feel they have more to lose than men by disclosing in order to be professionally successful, or emotional— women may feel more of a sense of helplessness and loss of control once information is disclosed.

We found surgery residents to have higher scores on the residency programme's Error Orientation Scale and the clinical unit's Managerial Safety Commitment Scale than medicine residents, but not on the Safety Culture Summary score. This is consistent with a previous survey of residents, which found that presentations of errors causing adverse events occurred 18% of the time in internal medicine ground rounds compared with 42% in surgery.[22] The differences between these two fields are likely due to divergent regulatory and cultural factors. Historically, morbidity and mortality rounds have served as a forum where surgeons learn from poor outcomes and aspire to identify their errors,[23] but this tradition is weaker in medicine training programmes.[24] The Accreditation Council for Graduate Medical Education (ACGME) requires that surgery morbidity and mortality conferences present and discuss 'all deaths and complications that occur on a weekly basis'. Historically, there has been no similar requirement for internal medicine.[24] Without a specific requirement to do so, adverse events and errors occurring in the medicine service may not be generally discussed.[21]

Several medicine residencies have developed programmes to address the current ACGME competency on Systems Based Practice,[13] by teaching systems-based thinking using root cause analysis of medical errors,[25] which require residents to develop an awareness of working in multidisciplinary teams to enhance patient safety, and participate in identifying system errors and implementing potential solutions.[13] Several studies have demonstrated the benefits of such educational

Table 4 Safety environment factors associated with disclosure and apology, by gender* and programme

	Team Psychological Safety Median range (0–33.3)	p Value†	Residency Programme Error Orientation Median range (0–33.3)	p Value†	Managerial Safety Commitment Median range (0–33.3)	p Value†	Safety Culture Summary Median range (0–100)	p Value†
Total	26.6		25.0		25.0		76.6	
Gender*		0.6		0.3		0.4		0.3
Men	27.8		25.0		26.4		79.1	
Women	26.6		24.4		25.0		76.0	
Programme		0.2		0.02		0.05		0.2
Medicine	27.8		23.9		23.6		75.2	
Surgery	26.6		26.1		27.8		80.5	
Disclosed		0.07		0.07		0.2		0.04
Yes	28.9		26.6		26.4		81.9	
No	26.6		24.4		25.0		76.0	
Apologised		1.0		0.05		0.01		0.05
Yes	27.8		25.5		26.4		79.6	
No	26.6		23.9		25.0		75.5	

*N=95; two respondents who did not enter gender were excluded from the analysis.
†p Values calculated using the Wilcoxon rank-sum test for medians.

interventions,[26–29] although to our knowledge, no study has attempted to measure changes in learning environment as a result of such interventions. Although such programmes are helpful, current ACGME guidelines do not require training programmes to address a thorough behavioural process of managing medical mistakes: accepting responsibility; discussing with colleagues; disclosing and apologising to patient; conducting an error analysis; and making changes in a practice setting designed to reduce future errors.[1 2 5 6 11 30]

The ability of residents to cope with medical error may be dependent on reassurance and learning opportunities provided by medical colleagues and supervisors.[26] Our findings would support this assertion, given the positive association between our derived Medical Culture Summary score and reporting of disclosure and apology. However the low frequency of disclosure and apology suggests that more work needs to be done within our training programmes to mitigate the negative effects of error to individuals, and gain potential benefits from more thorough processing of errors for individuals and the healthcare system.

There are several issues germane to housestaff and errors that are not addressed by our study. Some believe that an effective apology includes offering some form of reparation for the mistake.[31] We did not examine the issue of reparation. While the literature suggests that resident physicians who accept responsibility for their errors and discuss them are more likely to report other improvements in their medical practice,[2] we did not examine this phenomenon. Furthermore, while disclosure in a timely and appropriate manner may influence a patient's decision to pursue legal action,[1–3] we did not explore the relationship between our findings and legal action. In addition, though a doctor's emotional reaction to an error can last for years[32] and negative emotional responses are associated with increased odds of future self-perceived errors,[33] we did not evaluate emotional reactions or predict the future likelihood of error. Lastly, we did not directly examine the extent to which subjects were trained regarding coping with medical error, and hence could not determine if this training influenced their behaviour.

This study has several limitations. First, residents at only one academic medical centre were surveyed, so the results may not be generalisable. In addition, the residents were surveyed during 2004–2005, so it is possible that these results may no longer be accurate. During the past 6 years, the ACGME training requirements have increased their focus on systems-level thinking and training programmes have increasingly focused on reduction of error. However, the authors feel that the key findings of this study are relevant today. The rate of safety culture change is relatively slow, as demonstrated

Kronman AC, Paasche-Orlow M, Orlander JD. *BMJ Qual Saf* (2011). doi:10.1136/bmjqs-2011-000084

by a recent hospital survey administered by the Agency for Healthcare Research and Quality: average composite change in safety culture to change 1% over 1–2 years.[20] An increased focus on reduction in error does not translate directly to increases in individual accountability, apology or disclosure of error. Second, the modest sample size, limits opportunity for multivariable analyses as well as statistical power to detect potential associations. Third, our survey directed residents to consider a single error. We did this to focus respondents' attention on the details of an event that they would remember clearly to gain insight into aspects of organisational culture. This specific error may not be representative of most errors. In addition, as most of these errors occurred during the residents' first year of training, the expectation for disclosure and apology may be different than for the other years of training. However, a prior study that included trainees at our institution suggested that the responsibility of delivering bad news often falls to junior members of the team, including first-year trainees and medical students.[34] Fourth, the scales of organisational culture we used have rarely been used in healthcare settings. As such, the clinical significance of our observed score differences are unclear.

Despite the limitations, we successfully adapted survey tools previously used in a business environment to measure aspects of the learning environment of clinical trainees which are associated with disclosure and apology for medical error. This instrument needs to be validated in other institutions before proving its value as a metric in residency programme safety culture. If validated, such an instrument could be a valuable tool to assess changes in learning environments. Measuring culture and providing such feedback to leadership and staff is one of the safe practices recommended by the National Quality Forum to promote patient safety and reduce medical error.[35]

Measuring culture change requires a multimodal approach, of which this instrument could make a valuable contribution.[36] Since the ability to measure medical culture, and changes to it, is immature,[37] our study provides baseline measurements to help move the field further along. Developing learning environment metrics will be valuable to other institutions and training programmes in the coming years, as incremental programmatic changes in systems-level thinking and disclosure of medical error continue to impact the learning environments of residents.

Our results suggest a need for training programmes to provide trainees with structured, meaningful ways to cope with errors to prevent negative emotional responses, as well as create learning environments that facilitate disclosure of errors. Attention may need to be paid to explicate potential gender-related differences. All this is particularly important if, as a profession, we are to instil proper values, attitudes and responses to the inevitable occurrence of error in the next generation of physicians. As residency programmes incorporate systems-level thinking into residency education for patient safety and error prevention, it is important not to neglect the humanistic and interpersonal consequences of error for providers and patients. In order to do so, we need to develop measurement tools for learning environments. Further research is needed to identify successful environmental attributes that promote disclosure and healthy processing of medical errors.

Acknowledgements The authors wish to thank Victoria Brower, MPH, for creating and maintaining the study database, and Ignacio De La Cruz for help with manuscript preparation.

Funding This work was supported by an American Cancer Society Physician Training Award in Preventive Medicine (PT APM 97 185 04, Jane Liebshutz, Principal Investigator) and a Boston University Medical Center Risk Management Grant (Andrea Kronman, Principal Investigator).

Competing interests None.

Ethics approval Boston University Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. *To Err is Human: Building a Safer Health System*. Washington, DC: National Academy Press, 1999.
2. Wu AW, Folkman S, McPhee SJ, *et al*. Do house officers learn from their mistakes? *JAMA* 1991;265:2089–94.
3. Barach P, Small S. Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems. *Br Med J* 2000;320:759–63.
4. Leape LL. Error in medicine *JAMA* 1994;272:1851–7.
5. Goldberg RM. Coping with medical mistakes and errors in judgement. *Ann Emerg Med* 2002;39:287–92.
6. Christensen JF, Levinson W, Dunn PM. The heart of darkness: the impact of perceived mistakes on physicians. *J Gen Intern Med* 1992;7:424–31.
7. Mizrahi T. Managing medical mistakes: idealogy, insularity, and accountability among internists-in-training. *Soc Sci Med* 1984;19:135–46.
8. White AA, Gallagher TH, Krauss MJ, *et al*. The attitudes and experiences of trainees regarding disclosing medical errors to patients. *Acad Med* 2008;83:250–6.
9. Padmore JS, Jaeger J, Riesenberg LA, *et al*. 'Renters' or 'owners'? Residents' perceptions and behaviors regarding error reduction in teaching hospitals: a literature review. *Acad Med* 2009;84:1765–74.
10. Weissman JS, Annas CL, Epstein AM, *et al*. Error reporting and disclosure systems: views from hospital leaders. *JAMA* 2005;293:1359–66.
11. Hilfiker D. Facing our mistakes. *N Engl J Med* 1984;310:118–22.
12. Hoff T, Jameson L, Hannan E, *et al*. A review of the literature examining linkages between organizational factors, medical errors, and patient safety. *Med Care Res Rev* 2004;61:3–37.
13. Hoff TJ, Pohl H, Bartfield J. Creating a learning environment to produce competent residents: the roles of culture and context. *Acad Med* 2004;79:532–9.
14. Don H, Slocum JW Jr. Organizational climate: measures, research and contingencies. *Acad Manag J* 1974;17:255–80.
15. Edmondson A. Psychological safety and learning behavior in work teams. *Adm Sci Q* 1999;44:350–83.
16. Rybowiak V, Garst H, Frese M, *et al*. Error Orientation Questionnaire (EOQ): reliability, validity, and different language equivalence. *J Organ Behav* 1999;20:527–47.
17. Pronovost P, Weast B, Holzmueller C, *et al*. Evaluation of the culture of safety: survey of clinicians and managers in an academic medical center. *Qual Saf Health Care* 2003;12:405–10.

18. Fein SP, Hilborne HL, Spiritus EM, *et al*. The many faces of error disclosure: a common set of elements and a definition. *J Gen Intern Med* 2007;22:755–61.
19. Coffey M, Thomson K, Tallett S, *et al*. Pediatric residents' decision-making around disclosing and reporting adverse events: the importance of social context. *Acad Med* 2010;85:1619–25.
20. Sorra J, Famolaro T, Dyer N, *et al*. Hospital survey on patient safety culture: 2011 user comparative database report. AHRQ Publication No. 11-0030, April 2011. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/qual/hospsurvey11/. Accessed 19 Sep 2011.
21. Kaldjian LC, Jones EW, Rosenthal GE, *et al*. An empirically derived taxonomy of factors affecting physicians' willingness to disclose medical errors. *J Gen Intern Med* 2006;21:942–8.
22. Pierluissi E, Fischer MA, Campbell AR, *et al*. Discussion of medical errors in morbidity and mortality conferences. *JAMA* 2003;290:2838–42.
23. Gore DC. National survey of surgical morbidity and mortality conferences. *Am J Surg* 2006;191:708–14.
24. Orlander JD, Fincke BG. Morbidity and mortality conference: a survey of academic internal medicine departments. *J Gen Intern Med* 2003;18:656–8.
25. Voss JD, May NB, Schorling JB, *et al*. Changing conversations: teaching safety and quality in residency training. *Acad Med* 2008;83:1080–7.
26. Engel KG, Rosenthal M, Sutcliffe KM. Residents' responses to medical error: coping, learning, and change. *Acad Med* 2006;81:86–93.
27. Kroll L, Singleton A, Collier J, *et al*. Learning not to take it seriously: junior doctors' accounts of error. *Med Educ* 2008;42:982–90.
28. Stroud L, McIlroy J, Levinson W. Skills of internal medicine residents in disclosing medical errors: a study using standardized patients. *Acad Med* 2009;84:1803–8.
29. Coyle YM, Mercer SQ, Murphy-Cullen CL, *et al*. Effectiveness of a graduate medical education program for improving medical event reporting attitude and behavior. *Qual Saf Health Care* 2005;14:383–8.
30. Leape LL. Reporting of adverse events. *N Engl J Med* 2002;347:1633–8.
31. Lazare A. Apology in medical practice: an emerging clinical skill. *JAMA* 2006;296:1401–4.
32. Waterman AD, Garbutt J, Hazel E, *et al*. The emotional impact of medical errors on practicing physicians in the United States and Canada. *Jt Comm J Qual Patient Saf* 2007;33:467–76.
33. West CP, Huschka M, Novotny PJ, *et al*. Association of perceived medical errors with resident distress and empathy: a prospective longitudinal study. *JAMA* 2006;296:1071–8.
34. Orlander JD, Graeme Fincke B, Hermanns D, *et al*. Medical residents' first clearly remembered experiences of giving bad news. *J Gen Intern Med* 2002;17:825–40.
35. National Quality Forum (NQF). Safe practices for better healthcare—2009 update: a consensus report. Washington, DC: NQF;2009. ISBN 978-1-933875-23-1.
36. Measuring Culture Change: A Literature Review. 2006. Colorado Foundation for Medical Care (CFMC). Publication Number: PM-411-114 CO 2006. http://www.cfmc.org/files/nh/MCC%20Lit%20Review.pdf. Accessed 19 Sep 2011.
37. Paine LA, Rosenstein BJ, Sexton JB, *et al*. Assessing and improving safety culture throughout an academic medical centre: a prospective cohort study. *Qual Saf Health Care* 2010;19:547–54.

Journal of the American College of Cardiology
© 2011 by the American College of Cardiology Foundation
Published by Elsevier Inc.

Vol. 58, No. 6, 2011
ISSN 0735-1097/$36.00
doi:10.1016/j.jacc.2011.03.042

Biomarkers

Influence of Sex and Hormone Status on Circulating Natriuretic Peptides

Carolyn S. P. Lam, MBBS, MS,*‖¶ Susan Cheng, MD,*† Karen Choong, MD,‡ Martin G. Larson, ScD,*§ Joanne M. Murabito, MD, ScM,* Christopher Newton-Cheh, MD, MPH,*† Shalender Bhasin, MD,‡ Elizabeth L. McCabe, MS,*† Karen K. Miller, MD,† Margaret M. Redfield, MD,‖ Ramachandran S. Vasan, MD,*¶ Andrea D. Coviello, MD, MSE,*‡¶ Thomas J. Wang, MD*†

Framingham and Boston, Massachusetts; and Rochester, Minnesota

Objectives The aim of this study was to assess the relationship between sex hormones and natriuretic peptide levels in community-based adults.

Background Women have higher circulating natriuretic peptide concentrations than men, but the mechanisms for these sex-related differences and the impact of hormone therapy are unclear. Experimental studies suggest that androgens may suppress natriuretic peptide secretion.

Methods We measured N-terminal pro–B-type natriuretic peptide (NT-proBNP), total testosterone, and sex hormone–binding globulin plasma levels in 4,056 men and women (mean age 40 ± 9 years) from the Framingham Heart Study Third-Generation cohort. Sex/hormone status was grouped as: 1) men; 2) post-menopausal women not receiving hormone replacement therapy; 3) pre-menopausal women not receiving hormonal contraceptives; 4) post-menopausal women receiving hormone replacement therapy; and 5) pre-menopausal women receiving hormonal contraceptives.

Results Circulating NT-proBNP levels were associated with sex/hormone status (overall $p < 0.0001$). Men had lower NT-proBNP levels than women of all menopause or hormone groups, and women receiving hormonal contraceptives had higher NT-proBNP levels than women who were not receiving hormone therapy (all $p < 0.0001$). These relationships remained significant after adjusting for age, body mass index, and cardiovascular risk factors. Across sex/hormone status groups, free testosterone (FT) levels decreased and sex hormone–binding globulin levels increased in tandem with increasing NT-proBNP levels. In sex-specific analyses, NT-proBNP levels decreased across increasing quartiles of FT in men (p for trend <0.01) and women (p for trend <0.0001). Adjustment for FT markedly attenuated the association between sex/hormone status and NT-proBNP concentrations.

Conclusions These findings suggest that lower levels of circulating androgens and the potentiating effect of exogenous female hormone therapy contribute to the higher circulating NT-proBNP concentrations in women. (J Am Coll Cardiol 2011;58:618–26) © 2011 by the American College of Cardiology Foundation

The importance of understanding the effects of sex and hormone therapy on the cardiovascular system is underscored by the pronounced sex differences in the prevalence of cardiovascular disease, the increase in cardiovascular events in women following menopause, and concerns regarding the safety of hormone replacement therapy (HRT) in post-menopausal women (1–3). Studies of circulating cardiovascular biomarkers, such as the natriuretic peptides (NPs) (4–8), may provide a biological basis to better understand these sex-related differences in cardiovascular risk. The NPs exert hormonal (vasodilation, natriuresis, and

See page 627

*From the Framingham Heart Study, Framingham, Massachusetts; †Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts; ‡Section of Endocrinology, Diabetes, and Nutrition, Boston University School of Medicine and Boston Medical Center, Boston, Massachusetts; §Department of Mathematics and Statistics, Boston University, Boston, Massachusetts; ‖Mayo Clinic, Rochester, Minnesota; and the ¶Preventive Medicine and Cardiology Sections, Boston University School of Medicine, Boston, Massachusetts. This study is from the Framingham Heart Study of the National Heart, Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine, and is supported by NIH/NHLBI NO1-HC-25195, R01-HL-086875, and R01-HL-094755. Dr. Newton-Cheh serves on a scientific advisory board at Merck. Dr. Redfield is supported by HL-84907. All other authors have reported that they have no relationships to disclose. Drs. Coviello and Wang contributed equally to this work.

Manuscript received November 1, 2010; revised manuscript received March 15, 2011, accepted March 21, 2011.

aldosterone and endothelin suppression) and autocrine/paracrine (antihypertrophic, antifibrotic, and proangiogenic) protective cardiovascular effects. Sex is one of the strongest determinants of concentrations of circulating B-type natriuretic peptide (BNP) or N-terminal pro–B-type natriuretic peptide (NT-proBNP) in population-based studies, with women having consistently higher circulating levels than men (4–7). Nonetheless, the mechanisms underlying the sex-based difference in circulating NPs have not been established. Further, the effect of menstrual status, menopausal status, or exogenous hormone therapy on plasma NP concentrations in women remains unclear (6,8–10).

Previous studies have suggested that sex hormones play an important role in the regulation of NPs (11). On one hand, estrogens have been shown to exert a stimulating effect on the NP system (9,12) and are therefore postulated to mediate the "NP excess" in women compared with men. On the other hand, recent experimental (13,14) and cross-sectional human data (8,15) suggest that androgens may exert an inhibitory effect on the NP system, thus accounting for the lower NP levels in men and potentially explaining the lack of cardiovascular protection in men compared with women (8). However, prior clinical studies of the influence of androgens on NPs were limited to studying women (8,16) or children (15) and did not fully characterize both endogenous (menstrual phase) and exogenous (hormone therapy) variations. Further, previous investigations did not include measures of insulin resistance, which is a potential confounder because hyperinsulinemia is known to be associated with both sex hormones (hyperandrogenemia) and lowered NP concentrations (17–20).

We aimed to test the hypothesis that free testosterone (FT) is an important determinant of the relationship between sex/hormone status and circulating NPs in adults from the general population. Specifically, we hypothesized that higher FT levels in men compared with women and in women not receiving hormone therapy compared with those receiving hormone therapy may be related to lower circulating concentrations of NPs. To achieve our aim, we measured levels of testosterone and its primary binding protein, NP levels, and insulin measures in a large, community-based sample of predominantly middle-aged adults grouped by sex, menstrual status, and the presence or absence of hormone therapy.

Methods

Study sample. The Framingham Heart Study (21) is a community-based cohort investigation that began in 1948 with the recruitment of the Original Cohort, recruited a second generation in 1971 consisting of offspring of the Original Cohort and their spouses (Offspring Cohort), and most recently added a third generation in 2002 consisting of children of the Offspring Cohort (Third-Generation Cohort). Participants of the youngest cohort (Third-Generation Cohort), in whom plasma NT-proBNP concentrations were measured, were eligible for the current study. Those with prevalent heart failure (prior diagnosis of heart failure based on Framingham criteria [22]), myocardial infarction, or serum creatinine >2 mg/dl were excluded from the present investigation (n = 26). The index examinations took place during April 2002 to July 2005. All women had menstrual histories recorded by trained physicians using standardized questionnaires, including details regarding reproductive history, first day of the last menstrual period, frequency and regularity of menstrual cycles, and current usage of any hormone therapy. Participants were instructed to bring all of their current medications (taken within 1 month) with them to the examination, allowing verification by the examining physician. Anthropometric, blood pressure, and cardiovascular risk factor data were obtained at the index visit using standard protocols. All participants provided written informed consent, and protocols were approved by the Boston University Medical Center Institutional Review Board.

Abbreviations and Acronyms

BNP = B-type natriuretic peptide
FT = free testosterone
HC = hormonal contraceptives
HOMA = homeostatic model assessment
HRT = hormone replacement therapy
IR = insulin resistance
NP = natriuretic peptide
NT-proBNP = N-terminal pro–B-type natriuretic peptide
SHBG = sex hormone–binding globulin

Classification of sex/hormone status. Menopause was defined as the cessation of periods (in the absence of pregnancy) for at least 1 year because of naturally occurring (as opposed to surgical or medical) causes. Among postmenopausal women, those receiving HRT were defined by self-reported current usage of female hormone therapy, as well as validation by direct assessment of all of the participants' current medications by the examining physician. Among pre-menopausal women, those receiving hormonal contraceptives (HC) were defined by self-reported current usage of hormone-containing oral pills, injections, or implants for birth control or medical indications, as well as validation by direct assessment of all of the participants' current medications by the examining physician. Based on these definitions, sex/hormone status was classified into 5 categories (Fig. 1):

- Men
- Post-menopausal women not receiving HRT
- Pre-menopausal women not receiving HC
- Post-menopausal women receiving HRT currently
- Pre-menopausal women receiving HC currently

In secondary analyses, we estimated the menstrual phase (follicular vs. luteal vs. midcycle) at the index visit among pre-menopausal women with regular ongoing menstrual cycles (cycle lengths 28 to 30 days) who were not receiving HC. This determination was based on the fact that the interval from ovulation to menstruation (luteal phase) is fixed at 14 days, whereas the interval from the start of menstruation to ovulation (follicular phase) can vary de-

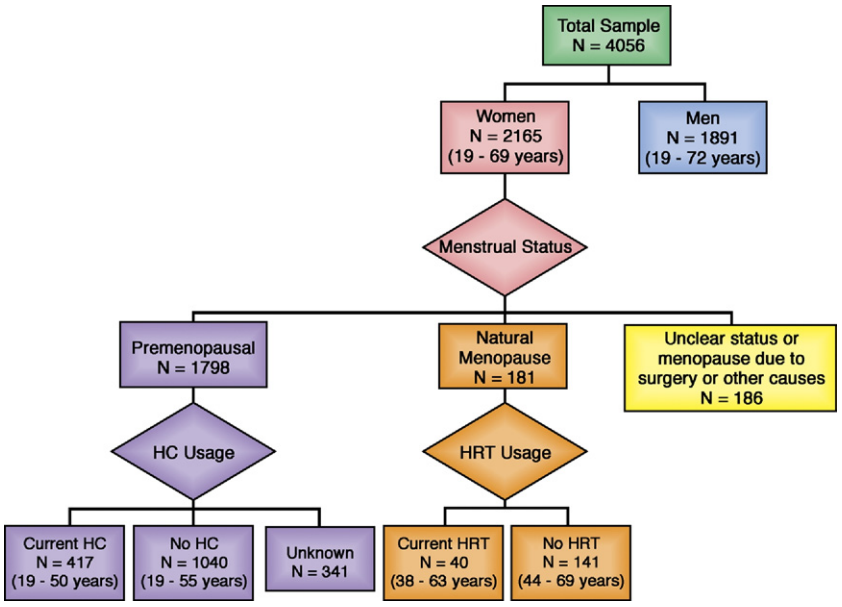


Figure 1 **Classification of Sex/Hormone Status**

The flow chart illustrates the classification of sex/hormone status into 1) men; 2) post-menopausal women not receiving hormone replacement therapy (HRT); 3) pre-menopausal women not receiving hormonal contraceptives (HC); 4) current HRT users; and 5) current HC users. The age range of participants in each group is indicated in brackets.

pending on cycle length, as previously established (23) and widely applied in similar studies (24–27). Thus, using the date of the examination, date of onset of the last menstrual period, and menstrual cycle length (specifically ascertained at the index examination), we derived the duration of the follicular phase in each regularly cycling woman, allowing 4 days between the follicular and luteal phases for the mid-cycle phase. The calculated durations were then used to assess whether NT-proBNP measurement on the date of the index visit occurred in the follicular, midcycle, or luteal phase of the menstrual cycle.

Measurement of circulating NT-proBNP, testosterone, and SHBG. Venous blood was drawn under fasting conditions between 8:00 and 9:00 AM. Samples were immediately stored at −70°C and analyzed in batches in 2009, allowing minimization of interassay variability and effects of temporal drift in the laboratory measurements. Plasma NT-proBNP levels were measured using a standard immunoassay (Roche Diagnostics, Indianapolis, Indiana), with a measurement range of 5 to more than 35,000 ng/l and intra-assay coefficient of variation of 2.7%. Serum total testosterone concentrations were quantified using a validated liquid chromatography–tandem mass spectrometry assay, with a lower detection limit of 2 ng/dl. Sex hormone–binding globulin (SHBG) concentrations were measured using an immunofluorometric assay (DELFIA-Wallac, Inc., Turku, Finland). Total testosterone concentrations are influenced by multiple factors, such as obesity in men and oral contraceptive therapy in women, and may not reflect androgen activity. We thus calculated FT using the law of mass action equation (28,29) and used FT in subsequent analyses.

Measurement of insulin resistance. Insulin resistance (IR) was calculated using the homeostatic model assessment (HOMA) equation: HOMA-IR = (FPG × FPI)/22.5, for which FPG = fasting plasma glucose (mmol/l) and FPI = fasting plasma insulin (mU/l) measured by enzyme-linked immunosorbent assay (Millipore Corporation, Billerica, Massachusetts) (30).

Statistical analyses. The association between log NT-proBNP and sex/hormone status was assessed using general linear models followed by pairwise comparisons between sex/hormone status groups with Bonferroni correction. The model was adjusted for covariates known to influence NT-proBNP concentrations (age, body mass index, systolic blood pressure, diastolic blood pressure, serum creatinine, and presence of antihypertensive medications, diabetes mellitus, and current smoking). We similarly assessed the association of log FT and log SHBG with sex/hormone status. Of a total of 2,165 women (Fig. 1), menopause status was unclear or not due to natural causes in 186, HC usage was unknown in 341, and FT/SHBG measurements or clinical covariates were missing in a further 25, leaving 1,613 women in the final multivariable models for sex/hormone status.

To assess the association between log NT-proBNP and circulating androgens, we used sex-pooled and sex-specific general linear models, in which the dependent variable was log NT-proBNP and predictors were log FT or log SHBG (separately), hormone status (among women; categories 2 to

Table 1 Baseline Characteristics

	All (N = 4,056)	Men (n = 1,891)	Women (n = 2,165)
Age, yrs	40 ± 9	40 ± 9	40 ± 9
Body mass index, kg/m^2	26.9 ± 5.6	28.0 ± 4.7	26.0 ± 6.1
Systolic blood pressure, mm Hg	117 ± 14	121 ± 13	113 ± 14
Diastolic blood pressure, mm Hg	75 ± 10	78 ± 9	73 ± 9
Serum creatinine, mg/dl	0.79 ± 0.15	0.90 ± 0.13	0.70 ± 0.11
Hypertension	16	21	12
Antihypertensive medications	8	10	7
Diabetes mellitus	3	4	2
Smoking	17	18	16
NT-proBNP, ng/l	28.1 (14.1, 52.6)	16.2 (8.1, 28.8)	42.9 (25.7, 72.2)
Log NT-proBNP	3.3 ± 1.0	2.7 ± 0.9	3.7 ± 0.8
FT, pg/ml	5.3 (2.1, 114.5)	119.0 (95.0, 152.0)	2.2 (1.4, 3.2)
Total testosterone, ng/dl	49.6 (23.7, 600.2)	617.7 (487.5, 786.1)	24.6 (17.7, 34.3)
SHBG, nmol/l	55.1 (35.2, 96.1)	37.0 (26.7, 50.2)	89.7 (58.7, 132.7)
HOMA-IR	0.98 (0.70, 1.49)	1.01 (0.78, 1.66)	0.89 (0.65, 1.31)

Values are mean ± SD, %, or median (25th, 75th percentiles). To convert FT concentration to SI units (pmol/l), multiply values in pg/ml by 3.47; to convert total testosterone concentration to SI units (nmol/l), multiple values in ng/dl by 0.0347.

FT = free testosterone; HOMA-IR = homeostatic model assessment of insulin resistance; NT-proBNP = N-terminal pro–B-type natriuretic peptide; SHBG = sex hormone–binding globulin.

5 of sex/hormone status as defined above), and covariates (listed above). We tested for interaction between log FT or log SHBG and hormone status among women and adjusted models for HOMA-IR among participants without diabetes. To further assess the association among sex/hormone status, log NT-proBNP, and circulating androgen levels, we analyzed multivariable models (sex pooled and in women alone), in which the dependent variable was log NT-proBNP and predictors sequentially included clinical covariates (listed above), sex/hormone status groups (defined above), and log FT. We assessed the beta coefficients for each sex/hormone status group, compared with the referent group of pre-menopausal women receiving HC, before and after adding log FT to the model. We also compared the type I sum of squares with the type III sum of squares for the class variable sex/hormone status to assess the percent variability in NT-proBNP levels due to sex/hormone status before (type I sum of squares) and after (type III sum of squares) the addition of log FT to the model, adjusting for all other covariates. Of a total of 2,165 women and 1,891 men, FT measurements were available for 2,144 women and 1,880 men, whereas SHBG measurements were available for 2,147 women and 1,880 men. Of these, there were no missing clinical covariates (age, body mass index, systolic blood pressure, diastolic blood pressure, serum creatinine, antihypertensive medications, diabetes mellitus, and smoking) in 2,123 women with FT measurements, 2,126 women with SHBG measurements, and 1,871 men. In addition to clinical covariates, HOMA-IR measurements were also available for 1,857 women with FT measurements, 1,859 women with SHBG measurements, and 1,650 men.

In the subgroup of pre-menopausal women with regular menstrual cycles in the absence of HC, we compared log NT-proBNP concentrations among menstrual phases (follicular, midcycle, and luteal) using general linear models with Bonferroni correction in pairwise comparisons among phases. We similarly compared log FT and log SHBG concentrations among menstrual phase groups. Models were adjusted for covariates known to influence NT-proBNP concentrations (listed above).

All analyses were performed using SAS software. Statistical significance was determined at a p value of <0.05 (Bonferroni-adjusted significance level required $p < 0.05$ [number of possible comparisons] in cases of multiple comparisons).

Results

Baseline characteristics. Characteristics of the study sample (N = 4,056; mean age 40 years) are shown in Table 1. As expected in a predominantly middle-aged, community-based sample, the prevalence of cardiovascular risk factors was low, and the majority (83%) of women were pre-menopausal (Fig. 1). Among pre-menopausal women, 23% (n = 417) were currently receiving HC; among post-menopausal women, 21% (n = 40) were currently receiving HRT. HRT consisted predominantly of oral combination therapy with estrogen and progesterone (n = 34 [85%]). Of note, the age range of participants was relatively narrow, even among post-menopausal women.

Association between sex/hormone status and circulating NT-proBNP levels. Log NT-proBNP was strongly associated with sex/hormone status (age-adjusted $p < 0.0001$) (Fig. 2A), with the lowest concentrations in men and highest concentrations in pre-menopausal women receiving HC. These relationships remained unchanged after adjusting for age, body mass index, systolic blood pressure, diastolic blood pressure, serum creatinine, and presence of antihypertensive medications, diabetes mellitus, and smok-

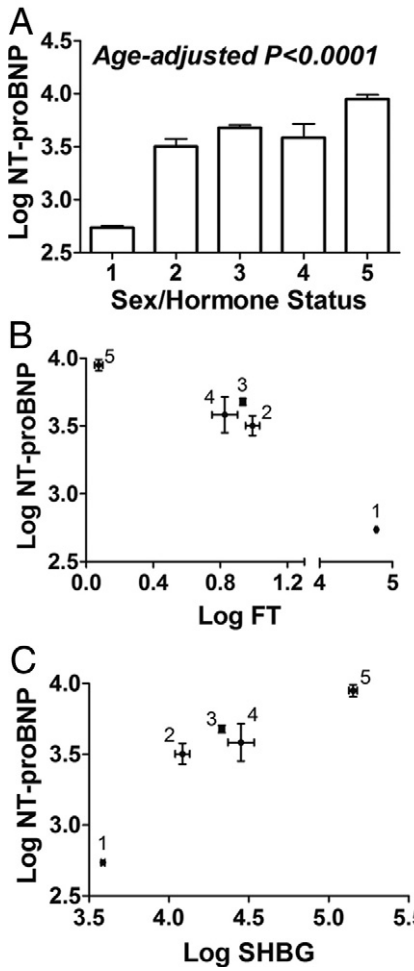


Figure 2 Association of Sex/Hormone Status With Circulating NT-proBNP, Testosterone, and SHBG

(A) Columns and **bars** represent the age-adjusted least squares means and SE of log N-terminal pro–B-type natriuretic peptide (NT-proBNP) by sex/hormone status (1 = men, 2 = post-menopausal women not receiving HRT, 3 = pre-menopausal women not receiving HC, 4 = post-menopausal HRT users, and 5 = pre-menopausal HC users). The p value shown is for the association between NT-proBNP and sex/hormone status overall; for pairwise comparisons within sex/hormone status groups, please refer to the text. **(B)** Bivariate plot showing the age-adjusted least squares means and SE of log free testosterone (FT) **(horizontal axis)** and log NT-proBNP **(vertical axis)** in each sex/hormone status group. **(C)** Bivariate plot showing the age-adjusted least squares means and SE of log sex hormone–binding globulin (SHBG; horizontal axis) and log NT-proBNP (vertical axis) in each sex/hormone status group. Abbreviations as in Figure 1.

ing. In pairwise comparisons, men had lower plasma NT-proBNP levels than women regardless of menopause status or hormone therapy (Bonferroni-corrected $p < 0.0001$). Among pre-menopausal women, levels of NT-proBNP were higher in those receiving HC (Bonferroni-corrected $p < 0.0001$). Among post-menopausal women, there was no difference in circulating NT-proBNP levels between current HRT users and nonusers (multivariable-adjusted $p = 0.39$). In additional analyses including an interaction term between year of blood collection and sex, there was no evidence for an interaction.

Association among circulating NT-proBNP, FT, and SHBG levels. As expected, men had the highest FT concentrations, whereas women receiving HC had the lowest FT and highest SHBG concentrations. Across sex/hormone status groups, FT levels decreased and SHBG levels increased in tandem with increasing NT-proBNP levels (Figs. 2B and 2C).

In sex-stratified analyses, NT-proBNP levels were lower in the highest quartiles of FT in both men (Fig. 3A) and women (Fig. 3B). Log NT-proBNP was inversely related to log FT and directly related to log SHBG in both men and women (Table 2). After adjustments for age, body mass index, systolic blood pressure, diastolic blood pressure, serum creatinine, antihypertensive medications, diabetes, and smoking, the association between log NT-proBNP and log FT was significant in women ($p < 0.0001$) but not men ($p = 0.78$). Each unit increase in log FT was associated with a 20% decrease in NT-proBNP levels among women, adjusting for clinical covariates. Results were similar after further adjustments for HOMA-IR among participants without diabetes. In both sexes, log NT-proBNP was related to log SHBG in multivariable analyses ($p < 0.0001$). Each unit increase in log SHBG was associated with a 19% increase in NT-proBNP levels among men and a 40% increase in NT-proBNP levels among women, adjusting for clinical covariates. There was no interaction between menopause or hormone status and log FT or log SHBG.

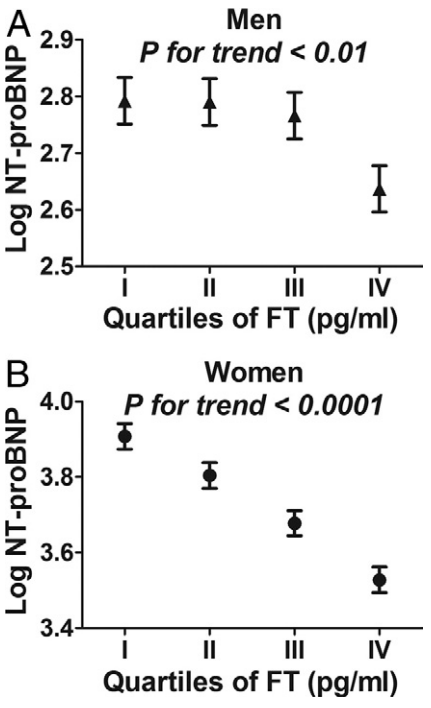


Figure 3 Association Between Circulating NT-proBNP and FT Levels

Least squares means and SE of log NT-proBNP are shown for each sex-specific quartile (I to IV) of FT in men **(triangles; A)** and women **(circles; B)**, respectively. Abbreviations as in Figure 2.

Table 2 Association Among Circulating NT-proBNP, Testosterone, and SHBG

	Men		Women	
Association With Log NT-proBNP	Beta* (SE)	p Value	Beta* (SE)	p Value
Entire sample (single variable)				
Log FT	−0.30 (0.11)	0.007	−0.54 (0.05)	<0.0001
Log SHBG	0.23 (0.03)	<0.0001	0.23 (0.02)	<0.0001
Entire sample (multivariable)†				
Log FT	−0.03 (0.11)	0.78	−0.47 (0.06)	<0.0001
Log SHBG	0.12 (0.03)	<0.0001	0.25 (0.02)	<0.0001
Nondiabetic patients only (multivariable)‡				
Log FT	−0.05 (0.12)	0.65	−0.45 (0.06)	<0.0001
Log SHBG	0.13 (0.03)	0.0001	0.24 (0.02)	<0.0001

*Beta coefficients represent the mean change in log NT-proBNP per 1 SD difference in log FT or log SHBG; For single variable analyses, n = 1,880 in men (both models), 2,144 in women (model containing log FT), and 2,147 in women (model containing log SHBG). †Adjusted for age, body mass index, systolic blood pressure, diastolic blood pressure, serum creatinine, antihypertensive medications, diabetes mellitus, and smoking; n = 1,871 in men (both models), 2,123 in women (model containing log FT), and 2,126 in women (model containing log SHBG). ‡Adjusted for age, body mass index, systolic blood pressure, diastolic blood pressure, serum creatinine, antihypertensive medications, smoking, and HOMA-IR; n = 1,650 in men (both models), 1,857 in women (model containing log FT), and 1,859 in women (model containing log SHBG).

Abbreviations as in Table 1.

In the entire sample, sex/hormone status explained 38% of the variability in NT-proBNP levels in multivariable models adjusted for age, body mass index, systolic blood pressure, diastolic blood pressure, serum creatinine, and presence of antihypertensive medications, diabetes mellitus, and smoking. This effect was larger than the contribution of any other clinical covariate. The addition of log FT to multivariable models adjusted for age, body size, and cardiovascular risk factors led to attenuation of the differences in log NT-proBNP among sex/hormone status groups (Table 3). After log FT was added to the multivariable model, clinical sex/hormone status only explained 1% of the variability of NT-proBNP levels.

In women alone, the addition of log FT to multivariable models similarly led to attenuation of the differences in log NT-proBNP levels among menopause or hormone therapy groups (Table 3). Menopause and hormone status explained 14% of the variability in NT-proBNP levels in women and in multivariable models adjusted for clinical covariates but only 2% after log FT was added to the multivariable model. Thus, the vast majority of the variability in NT-proBNP due to menopause and hormone status in women appeared to be attributable to differences in FT concentrations.

In analyses restricted to participants not receiving any antihypertensive therapy, the associations among sex/hormone status, NT-proBNP, and FT were unchanged (not shown).

Subgroup analysis by menstrual phase. Among 546 premenopausal women with regular menstrual cycles in the absence of HC (Table 4), NT-proBNP levels were lower in the midcycle phase than in the follicular or luteal phase ($p = 0.014$ for midcycle vs. follicular phase; $p = 0.015$ for midcycle vs. luteal phase; Bonferroni-corrected $p < 0.05$ for

Table 3 Multivariable Models Assessing the Association Among Sex/Hormone Status, Androgens, and NT-proBNP Concentrations

	Model Without FT		Model With FT	
Association With Log NT-proBNP	Beta* (SE)	p Value	Beta* (SE)	p Value
Entire sample (N = 3,484)				
Men	−1.12 (0.05)	<0.0001	−0.51 (0.14)	0.0004
Post-menopausal women without HRT	−0.46 (0.09)	<0.0001	−0.35 (0.09)	0.0001
Pre-menopausal women without HC	−0.25 (0.05)	<0.0001	−0.14 (0.05)	0.008
Post-menopausal women with HRT	−0.34 (0.14)	0.015	−0.24 (0.14)	0.083
Pre-menopausal women with HC	Referent	Referent	Referent	Referent
Log FT	—	—	−0.27 (0.06)	<0.0001
Women only (n = 1,613)				
Post-menopausal women without HRT	−0.35 (0.09)	0.0001	−0.17 (0.10)	0.073
Pre-menopausal women without HC	−0.22 (0.05)	<0.0001	−0.05 (0.06)	0.385
Post-menopausal women with HRT	−0.25 (0.14)	0.069	−0.09 (0.14)	0.499
Pre-menopausal women with HC	Referent	Referent	Referent	Referent
Log FT	—	—	−0.43 (0.07)	<0.0001

*Beta coefficients represent the mean difference in log NT-proBNP in the corresponding sex/hormone status group compared with that in pre-menopausal women receiving hormonal contraceptives (HC) (referent group) or mean change in log NT-proBNP per 1 SD difference in log FT, adjusted for age, body mass index, systolic blood pressure, diastolic blood pressure, serum creatinine, antihypertensive medications, diabetes mellitus, and smoking.

HC = hormonal contraceptives; HRT = hormone replacement therapy; other abbreviations as in Table 1.

Table 4 Subgroup Analysis by Menstrual Phase

Multivariable-Adjusted* Least-Squares Mean (SE)	Menstrual Phase			p Value		
	Follicular (n = 262)	Midcycle (n = 70)	Luteal (n = 214)	Follicular vs. Midcycle	Follicular vs. Luteal	Luteal vs. Midcycle
Log NT-proBNP	3.71 (0.05)	3.45 (0.09)	3.71 (0.05)	0.014	0.97	0.015
Log FT	0.91 (0.03)	1.05 (0.06)	0.93 (0.04)	0.045	0.642	0.100
Log SHBG	4.33 (0.03)	4.38 (0.05)	4.33 (0.03)	0.348	0.943	0.386

*Adjusted for age, body mass index, systolic blood pressure, diastolic blood pressure, serum creatinine, antihypertensive medications, diabetes mellitus, and smoking.
Abbreviations as in Table 1.

both) in multivariable models. There was a corresponding trend toward higher FT levels in the midcycle phase compared with that in the follicular or luteal phase ($p = 0.045$ for midcycle vs. follicular phase; $p = 0.099$ for midcycle vs. luteal phase). When log FT was added to multivariable models, the association between menstrual phase and log NT-proBNP became nonsignificant. There was no association between SHBG levels and menstrual phase.

Discussion

Principal findings. In a large sample of men and women from the general community, sex and exogenous hormone therapy were the largest determinants of variation in circulating NT-proBNP levels. Men had lower plasma NT-proBNP concentrations than women regardless of menopause status or hormone therapy, whereas women receiving HC had higher NT-proBNP concentrations than women without hormone therapy. Among sex/hormone status groups, men also had the highest FT and lowest SHBG concentrations; conversely, women receiving HC had the lowest FT and highest SHBG concentrations. In both sexes, increasing NT-proBNP levels were related to decreasing FT and increasing SHBG concentrations. Accounting for FT greatly attenuated the differences in circulating NT-proBNP levels among sex/hormone status groups, even after adjustments for known clinical covariates and restricting analyses to women. These findings are consistent with the hypothesis that androgens suppress NT-proBNP and suggest that differences in FT concentrations may largely explain the sex- and hormone-related differences in circulating NPs.

Androgen regulation of NP concentrations. Several lines of evidence suggest that testosterone may exert a suppressive effect on the NP system and thus mediate a "BNP deficiency" in men compared with women. In male rats, orchiectomy produced marked increases in plasma NP levels, and testosterone replacement restored NP concentrations to baseline (14). In isolated perfused rat atria, testosterone suppressed volume-stimulated release of atrial NP (13). However, contradictory results have also been reported regarding the effect of testosterone on atrial NP gene expression (31) and synthesis in cultured rat myocytes (32). In clinical studies, inverse correlations between BNP and FT levels have been observed in women in the Dallas Heart Study (8) and in a small Japanese study (16). Androgen receptor antagonism for prostate cancer has been associated with large increases in levels of NT-proBNP (33). In male children and adolescents, associations of decreasing FT, increasing SHBG, and increasing NT-proBNP levels have been noted (15).

Our findings extend prior observations to a large community-based cohort of middle-aged men and women. The demonstrated association between NT-proBNP and androgens offers a potential unifying explanation for sex and hormonal status-related differences in NP concentrations. In men, low BNP concentrations may be related to the suppressive effects of high concentrations of FT. A nonlinear relationship with NT-proBNP at such high concentrations of FT may explain the failure to detect a statistically significant relationship between FT and NT-proBNP in men following multivariable adjustment. In women, FT concentrations are more than 50-fold lower than that in men and are exquisitely sensitive to SHBG concentrations.

Estrogen regulation of NP concentrations. Previous studies have also shown that estrogens may exert a stimulatory effect on the NP system (11). In female rats, pretreatment with estradiol and progesterone stimulated atrial NP gene expression (31). In post-menopausal women, administration of estrogens produced a rise in plasma levels of BNP (9). A complex interplay of factors has been postulated, in which estrogens modulate NP production via known effects on the renin-angiotensin system (12).

Measurements of estradiol and estrone were not available for the current study, limiting our conclusions regarding the role of estrogens in mediating the observed difference in NT-proBNP levels. However, according to the hypothesis of a stimulatory effect of estrogens on NPs, we would have expected post-menopausal women and men to have similar NT-proBNP levels given their similar estrogen levels, yet NT-proBNP concentrations were much higher in the former. Similarly, based on higher endogenous estrogens in pre-menopausal than post-menopausal women, we would have expected higher NT-proBNP levels in the former, but levels were similar in the 2 groups. Lastly, given that the midcycle (ovulatory) phase of the menstrual cycle is associated with higher estrogen levels compared with the follicular or luteal phases, higher NT-proBNP levels would be expected midcycle; however, levels were paradoxically lower in women at midcycle. Interestingly, each of the above observations can potentially be explained by variation in FT, which is highest in men, comparable in pre- and post-

menopausal women, and higher during the midcycle phase than the follicular or luteal phase. These findings are consistent with prior studies showing that circulating NT-proBNP concentrations are not correlated with measured estradiol concentrations in females (15,16).

Nonetheless, in the absence of direct measurements of estrogens, a role for estrogen-stimulated increases in NT-proBNP cannot be excluded. An explanation based purely on androgen suppressive effects would be inadequate to account for the known rapid fall in NP concentrations during the first year of life or the similar levels of NT-proBNP in male adolescents and male pre-pubertal children (34). In aggregate, it is likely that both the stimulatory effects of estrogens and inhibitory effects of testosterone contribute to the regulation of BNP concentrations during the life course. The relative concentrations of these sex hormones may also be an important factor. Additional population-based studies including pre- and post-pubertal individuals of both sexes are warranted.

Circulating NPs and use of hormone therapies. Previous studies examining the impact of HRT on BNP in post-menopausal women have produced conflicting results (6,8–10), and none have examined the impact of HC on circulating BNP in pre-menopausal women. Our study indicates that usage of HC in pre-menopausal women is associated with higher circulating NT-proBNP levels compared with no usage. This may be due to direct stimulatory effects of estrogens on the NP system, a reduction in FT-mediated suppression of NT-proBNP secondary to increased SHBG from oral estrogens, or indirect effects of oral estrogens and progestins acting via the renin-angiotensin system to modulate NP levels (12). Consistent with previous studies (8,16), we did not detect a significant association between naturally occurring menopause and NT-proBNP concentrations after accounting for age, although the relatively small number of post-menopausal women in our middle-aged sample may have limited our statistical power to detect a difference. We were similarly unable to demonstrate a significant association between usage of HRT and circulating NT-proBNP, in contrast to findings from Olmsted County (6). This could be due to small numbers of HRT users in our current sample because we included younger women who were recruited following publication of the Women's Health Study and the nationwide reduction in HRT prescription rates (2,35). Of note, the effects of exogenous female sex hormone therapies are known to vary with the route of administration (36), formulation (37), and composition; for example, progestogens in HC may exert both androgenic and antiandrogenic effects (38). These details were not available in our study but represent important areas for future study.

Study strengths and limitations. Strengths of this study include the large sample size, community-based design, careful recording of menstrual history, standardized examinations with routine blood collection, and uniform ascertainment of cardiovascular risk factors including IR. Measurements of estrogens and details regarding the individual hormone components of HC were not available. The accurate detection of an ovulatory cycle or hormonal changes in reproductive aging in women requires specific measurements of female sex hormones and gonadotropins. A role for estrogen-stimulated increases in circulating BNP cannot be excluded based on these data. FT was not directly measured, but estimated FT concentrations from total testosterone (by mass spectrometry) and SHBG (by radioimmunoassay) correlate well with direct measurements by equilibrium dialysis (39–42). Biologically active atrial NP (43) and BNP (44) provide physiologically meaningful information but are less practical for measurement in large, ambulatory cohorts composed of predominantly healthy individuals, in part because of the high proportion of values censored by the detection limit of the mature NP assays (4,5). We acknowledge the potential for residual confounding by unmeasured comorbidities and their pharmacological therapies, as well as the limited ability to draw conclusions regarding causality from these observational data. Nonetheless, our findings are consistent with experimental data on the effects of testosterone on NPs.

Conclusions

Circulating NT-proBNP levels were related to sex and exogenous hormone therapy in men and women from the general community. Suppression of NPs by androgens may account for sex- and hormone-related differences in NT-proBNP concentrations. Given the known cardioprotective effects of BNP (45), further studies are warranted to elucidate how these mechanisms may contribute to the well-described sex-related differences in cardiovascular risk.

Reprint requests and correspondence: Dr. Carolyn S. P. Lam, National University Health System, Tower Block Level 9, 1E Kent Ridge Road, Singapore 119228. E-mail: carolyn_lam@nuhs.edu.sg.

REFERENCES

1. Grodstein F, Stampfer M. The epidemiology of coronary heart disease and estrogen replacement in postmenopausal women. Prog Cardiovasc Dis 1995;38:199–210.
2. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA 2002;288:321–33.
3. Beral V. Breast cancer and hormone-replacement therapy in the Million Women Study. Lancet 2003;362:419–27.
4. Wang TJ, Larson MG, Levy D, et al. Impact of age and sex on plasma natriuretic peptide levels in healthy adults. Am J Cardiol 2002;90:254–8.
5. Vasan RS, Benjamin EJ, Larson MG, et al. Plasma natriuretic peptides for community screening for left ventricular hypertrophy and systolic dysfunction: the Framingham Heart Study. JAMA 2002;288:1252–9.
6. Redfield MM, Rodeheffer RJ, Jacobsen SJ, Mahoney DW, Bailey KR, Burnett JC Jr. Plasma brain natriuretic peptide concentration: impact of age and gender. J Am Coll Cardiol 2002;40:976–82.
7. Abdullah SM, Khera A, Das SR, et al. Relation of coronary atherosclerosis determined by electron beam computed tomography and plasma levels of N-terminal pro-brain natriuretic peptide in a multiethnic population-based sample (the Dallas Heart Study). Am J Cardiol 2005;96:1284–9.

8. Chang AY, Abdullah SM, Jain T, et al. Associations among androgens, estrogens, and natriuretic peptides in young women: observations from the Dallas Heart Study. J Am Coll Cardiol 2007;49:109–16.
9. Maffei S, Del Ry S, Prontera C, Clerico A. Increase in circulating levels of cardiac natriuretic peptides after hormone replacement therapy in postmenopausal women. Clin Sci (Lond) 2001;101:447–53.
10. Kawano H, Nagayoshi Y, Soejima H, et al. B-type natriuretic peptide after hormone therapy in postmenopausal women with chest pain and normal coronary angiogram. Menopause 2008;15:352–6.
11. Clerico A, Fontana M, Vittorini S, Emdin M. The search for a pathophysiological link between gender, cardiac endocrine function, body mass regulation and cardiac mortality: proposal for a working hypothesis. Clin Chim Acta 2009;405:1–7.
12. Kuroski de Bold ML. Estrogen, natriuretic peptides and the renin-angiotensin system. Cardiovasc Res 1999;41:524–31.
13. Deng Y, Kaufman S. The influence of reproductive hormones on ANF release by rat atria. Life Sci 1993;53:689–96.
14. Hwu CM, Tsai SC, Lau CP, et al. Increased concentrations of atrial and plasma atrial natriuretic peptide in castrated male rats. Life Sci 1993;52:205–12.
15. Saenger AK, Dalenberg DA, Bryant SC, Grebe SK, Jaffe AS. Pediatric brain natriuretic peptide concentrations vary with age and sex and appear to be modulated by testosterone. Clin Chem 2009;55:1869–75.
16. Yasui T, Matsui S, Saijo A, et al. Association of N-terminal pro B-type natriuretic peptide and sex hormone-binding globulin in non-obese peri- and postmenopausal women. Clin Chim Acta 2010; 411:280–4.
17. Reinecke H, Bogdanski J, Woltering A, et al. Relation of serum levels of sex hormone binding globulin to coronary heart disease in postmenopausal women. Am J Cardiol 2002;90:364–8.
18. Hautanen A. Synthesis and regulation of sex hormone-binding globulin in obesity. Int J Obes Relat Metab Disord 2000;24 Suppl 2:S64–70.
19. Sutton-Tyrrell K, Wildman RP, Matthews KA, et al. Sex-hormone-binding globulin and the free androgen index are related to cardiovascular risk factors in multiethnic premenopausal and perimenopausal women enrolled in the Study of Women Across the Nation (SWAN). Circulation 2005;111:1242–9.
20. Wang TJ, Larson MG, Keyes MJ, Levy D, Benjamin EJ, Vasan RS. Association of plasma natriuretic peptide levels with metabolic risk factors in ambulatory individuals. Circulation 2007;115:1345–53.
21. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health 1951;41:279–81.
22. McKee PA, Castelli WP, McNamara PM, Kannel WB. The natural history of congestive heart failure: the Framingham study. N Engl J Med 1971;285:1441–6.
23. Norman R, Phillipson G. The normal menstrual cycle: changes throughout life. In: Fraser IS, Jansen RPS, Lobo RA, Whitehead MI, editors. Estrogens and Progestogens in Clinical Practice. New York, NY: Churchill Livingstone, 1998:105.
24. Lohstroh PN, Chen J, Ba J, et al. Bone resorption is affected by follicular phase length in female rotating shift workers. Environ Health Perspect 2003;111:618–22.
25. Reichman ME, Judd JT, Taylor PR, Nair PP, Jones DY, Campbell WS. Effect of dietary fat on length of the follicular phase of the menstrual cycle in a controlled diet setting. J Clin Endocrinol Metab 1992;74:1171–5.
26. Barnett JB, Woods MN, Lamon-Fava S, et al. Plasma lipid and lipoprotein levels during the follicular and luteal phases of the menstrual cycle. J Clin Endocrinol Metab 2004;89:776–82.
27. Kozlowski PA, Williams SB, Lynch RM, et al. Differential induction of mucosal and systemic antibody responses in women after nasal, rectal, or vaginal immunization: influence of the menstrual cycle. J Immunol 2002;169:566–74.
28. Mazer NA. A novel spreadsheet method for calculating the free serum concentrations of testosterone, dihydrotestosterone, estradiol, estrone and cortisol: with illustrative examples from male and female populations. Steroids 2009;74:512–9.
29. Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. J Clin Endocrinol Metab 1999;84:3666–72.
30. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 1985;28:412–9.
31. Hong M, Yan Q, Tao B, et al. Estradiol, progesterone and testosterone exposures affect the atrial natriuretic peptide gene expression in vivo in rats. Biol Chem Hoppe Seyler 1992;373:213–8.
32. Matsubara H, Hirata Y, Yoshimi H, et al. Effects of steroid and thyroid hormones on synthesis of atrial natriuretic peptide by cultured atrial myocytes of rat. Biochem Biophys Res Commun 1987;145:336–43.
33. Dockery F, Bulpitt CJ, Agarwal S, et al. Anti-androgens increase N-terminal pro-BNP levels in men with prostate cancer. Clin Endocrinol (Oxf) 2008;68:59–65.
34. Nir A, Lindinger A, Rauh M, et al. NT-pro-B-type natriuretic peptide in infants and children: reference values based on combined data from four studies. Pediatr Cardiol 2009;30:3–8.
35. Hersh AL, Stefanick ML, Stafford RS. National use of postmenopausal hormone therapy: annual trends and response to recent evidence. JAMA 2004;291:47–53.
36. Devineni D, Skee D, Vaccaro N, et al. Pharmacokinetics and pharmacodynamics of a transdermal contraceptive patch and an oral contraceptive. J Clin Pharmacol 2007;47:497–509.
37. Stanosz S, Zochowska E, Safranow K, Sieja K, Stanosz M. Influence of modified transdermal hormone replacement therapy on the concentrations of hormones, growth factors, and bone mineral density in women with osteopenia. Metabolism 2009;58:1–7.
38. Wiegratz I, Kuhl H. Metabolic and clinical effects of progestogens. Eur J Contracept Reprod Health Care 2006;11:153–61.
39. Miller KK, Rosner W, Lee H, et al. Measurement of free testosterone in normal women and women with androgen deficiency: comparison of methods. J Clin Endocrinol Metab 2004;89:525–33.
40. Wang C, Shiraishi S, Leung A, et al. Validation of a testosterone and dihydrotestosterone liquid chromatography tandem mass spectrometry assay: interference and comparison with established methods. Steroids 2008;73:1345–52.
41. Wang C, Catlin DH, Demers LM, Starcevic B, Swerdloff RS. Measurement of total serum testosterone in adult men: comparison of current laboratory methods versus liquid chromatography-tandem mass spectrometry. J Clin Endocrinol Metab 2004;89:534–43.
42. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. J Clin Endocrinol Metab 2007;92:405–13.
43. Clerico A, Recchia FA, Passino C, Emdin M. Cardiac endocrine function is an essential component of the homeostatic regulation network: physiological and clinical implications. Am J Physiol Heart Circ Physiol 2006;290:H17–29.
44. Clerico A, Emdin M. Diagnostic accuracy and prognostic relevance of the measurement of cardiac natriuretic peptides: a review. Clin Chem 2004;50:33–50.
45. Lee CY, Burnett JC Jr. Natriuretic peptides and therapeutic applications. Heart Fail Rev 2007;12:131–42.

Key Words: hormones ■ natriuretic peptides ■ sex.

HSR Health Services Research

DOI: 10.1111/j.1475-6773.2011.01283.x
METHODS ARTICLE

Comparative *Logic Modeling* for Policy Analysis: The Case of HIV Testing Policy Change at the Department of Veterans Affairs

Erika M. Langer, Allen L. Gifford, and Kee Chan

Objective. Logic models have been used to evaluate policy programs, plan projects, and allocate resources. *Logic Modeling* for policy analysis has been used rarely in health services research but can be helpful in evaluating the content and rationale of health policies. Comparative *Logic Modeling* is used here on human immunodeficiency virus (HIV) policy statements from the Department of Veterans Affairs (VA) and Centers for Disease Control and Prevention (CDC). We created visual representations of proposed HIV screening policy components in order to evaluate their structural logic and research-based justifications.

Data Sources and Study Design. We performed content analysis of VA and CDC HIV testing policy documents in a retrospective case study.

Data Collection. Using comparative *Logic Modeling*, we examined the content and primary sources of policy statements by the VA and CDC. We then quantified evidence-based causal inferences within each statement.

Principal Findings. VA HIV testing policy structure largely replicated that of the CDC guidelines. Despite similar design choices, chosen research citations did not overlap. The agencies used evidence to emphasize different components of the policies.

Conclusion. Comparative *Logic Modeling* can be used by health services researchers and policy analysts more generally to evaluate structural differences in health policies and to analyze research-based rationales used by policy makers.

Key Words. Evidence-based practice, HIV, health policy, Centers for Disease Control and Prevention (U.S.), Veterans Affairs (U.S.)

Health care decision makers have used research evidence to justify adoption of Centers for Disease Control and Prevention (CDC) human immunodeficiency virus (HIV) testing guidelines into a variety of health care settings. In 2006, the CDC released revised HIV testing guidelines for adults and adolescents (Branson et al. 2006). Within the U.S. Department of Veterans Affairs

(VA), reports emerged indicating that HIV testing rates were low (Owens et al. 2007), prompting reevaluation of VA HIV testing practices and policies. As a result, in 2009, the VA adopted CDC recommendations by eliminating written informed consent requirements for HIV testing and making testing a routine part of Veterans' health services (Department of Veterans Affairs 2009).

Translation of specific clinical recommendations and evidence into policies and practice within health care systems is a major challenge. If the structural logic and rationale of the CDC recommendations are not clear, initiatives and resource allocations necessary for a change in testing policy could be difficult to implement. James and Jorgensen (2009) have suggested that research utilization theory offers a robust conceptual framework for assessing the policy process. There is a pressing need to use effective evaluation tools to reveal the evidence-based resources, inputs, and outputs of policy in a systematic, logical, and transparent manner as health care settings implement the CDC guidelines. *Logic Modeling* is a technique that offers a way to analyze and quantify how policy makers use research—how their "research utilization" informs the process and outcomes of translating specific research-based knowledge into evidence-based practice (Rich 1997).

To understand research utilization, we propose a novel application of the logic model. *Logic Modeling* provides a visual representation of input, throughput, and output components that are brought together to produce intended change. While *Logic Modeling* can be used after completion of primary activities as a way of evaluating how well policy was able to meet intended outcomes, it may also be used earlier in institutional change processes to plan for and guide future evaluation. *Logic Modeling* as an evaluation tool has been used widely to examine the resources, inputs, outputs, and outcomes of programs in a clear and systematic fashion. Therefore, we applied

---

Address correspondence to Kee Chan, Ph.D., Assistant Professor, Department of Health Sciences, Boston University College of Health and Rehabilitation Sciences: Sargent College, 635 Commonwealth Avenue, Boston, MA 02215; e-mail: keechan@bu.edu. Erika M. Langer, M.S., and Allen L. Gifford, M.D., are with the Department of Health Policy and Management, Boston University School of Public Health, Boston, MA. Erika M. Langer, M.S., is also with Department of Health Sciences, Boston University College of Health and Rehabilitation Sciences: Sargent College, Boston, MA. Allen L. Gifford, M.D., is also with the Section of General Internal Medicine, Boston University School of Medicine, Boston, MA. Allen L. Gifford, M.D., and Kee Chan, Ph.D., are with the Center for Health Quality, Outcomes and Economic Research, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA. Kee Chan, Ph.D., is also with the Department of Epidemiology, Boston University School of Medicine, Boston, MA.

this to better understand the evolution of VA HIV testing and assess how evidence was brought to bear on the policy design.

Over the past 40 years, logic models have largely been applied to evaluations of specific social programs (Wholey 1987), but *Logic Modeling* has also been used in both community-based and systems-level initiatives (Julian 1997; Moyer, Verhovsek, and Wilson 1997; Kaplan and Garrett 2005), in planning (Dwyer 1996; Macaskill et al. 2000), management (Millar, Simone, and Carnevale 2001), and preevaluation to develop indicators and document outcomes (Kellogg Foundation 2004; Innovation Network 2008). *Logic Modeling* can stimulate reflection (Moyer, Verhovsek, and Wilson 1997), enable communication, and promote continued learning toward health service objectives (Kellogg Foundation 2004; Innovation Network 2008).

Policy documents such as the VA HIV directive and CDC HIV recommendations represent endpoints within the decision making process and are particularly relevant to successful replication of policies within other contexts. Used as primary sources for analysis, policy documents can counter the problematic "fuzziness" of health policy by providing insights into the formal rules and intentions of policy measures (Kroneman and van der Zee 1997). James and Jorgensen (2009) have suggested that by working backward from final policy statements, it is possible to reconstruct policy decision making to explore knowledge utilization, a term which encompasses both scientific and nonscientifically generated information sources apparent in the policy making process (Trostle, Bronfman, and Langer 1999; Dobrow, Goel, and Upshur 2004). In this formulation, "research utilization" is a sub-type of knowledge utilization that can describe the development of evidence-based practices.

Use of research data in policy making has previously been characterized as instrumental (i.e., direct, actionable), conceptual (i.e., diffuse, gradually enlightening), or symbolic (i.e., strategic, tactical) (Pelz 1978; Weiss 1979; Trostle, Bronfman, and Langer 1999; Hanney et al. 2003; Almeida and Báscolo 2006). We explore the symbolic use of research evidence. This use can occur to attain political legitimacy, support a position, give confidence, reduce uncertainty, and raise financial resources for policy decisions (Hanney et al. 2003).

Despite the personal, economic, and public health benefits of HIV screening, the revised 2006 CDC guidelines have not been universally adopted. Lack of consensus in how testing policy should be applied may stem in part from lack of transparency in how evidence-based resources, inputs, and outputs have been synthesized, translated, and adopted in different settings. We propose that when informed by research utilization theory, policy *Logic*

Modeling can generate this transparency. Modeling policy alongside causal assumptions allows us to compare across policy settings in order to better understand the rationales of specific policy choices. Where research-based justifications are entirely lacking in the model, there may exist gaps in CDC or VA policy logic. We propose that by generating insights into the policy making process, comparative *Logic Modeling* is thus an effective tool for health policy analysis.

MATERIALS AND METHODS

Using final CDC and VA policy statements on HIV testing (Branson et al. 2006; Department of Veterans Affairs 2009), this case study examines retrospectively the structural logic and research-based rationale of each policy. To describe the structural logic of each agency's policy design, we adopted the *outcomes-based logic model* from W. K. Kellogg Foundation (2004). We used content analysis of agency policy documents to identify elements of the five policy design components presented in the model. These components are resources/inputs, activities, outputs, outcomes, and impact (labeled 1–5 in Figure 1).

Comparison of Design Logic

Resources/inputs include space, technology, equipment, materials, and the human, financial, organizational, and community inputs available to direct toward proposed activities. *Activities* include processes, tools, events, technology, and actions needed to bring about intended change, and encompass the services, products, advocacy, and infrastructures of the intervention. *Outputs* are the direct measurable or tangible products of activities quantified as types, levels, and targets of delivered services. They are not themselves the anticipated change, but they help to assess how well change is being implemented. *Outcomes* are the individual, community, systematic, or organizational changes

Figure 1: Logic Model with Policy Design Components, 1–5, and Research-Based Links, A–D

Source: Adapted from Kellogg Foundation (2004).

to behavior, knowledge, skills, status or level of functioning. *Impact* is the fundamental change that occurs over the longer term (Kellogg Foundation 2004; Innovation Network 2008).

We used these definitions to construct separate logic models for the CDC and VA HIV testing policy statements. First, we categorized direct quotes from the statements according to the components of the logic model, and we cross checked this categorization between researchers. Next, we compared structural components across logic models to determine the extent of policy overlap, and we abbreviated these quotes for the purpose of display (see Tables 1 and 2). Sections of the CDC guidance pertaining to HIV screening for adolescents, pregnant women, and their infants were excluded from the analysis, as maternity care is typically provided outside the VA, and the agency does not provide care for adolescents or infants (Department of Veterans Affairs 2010).

Comparison of Design Rationales

Next, we adapted the logic model to draw specific attention to the presence or absence of a research-based rationale or evidence-based practice for policy design. The research-based policy rationale was described through use of the model arrows or "links" (labeled A–D in Figure 1). These represent causal inferences of the design which connect one policy component to another in an "if-then" statement (Kellogg Foundation 2004). The causal inferences of link B, for example, state that if the policy includes certain *activities*, then it will produce a particular *output*. This causal inference may or may not be supported explicitly by research evidence. While causal inferences may be supported by a variety of information types, our study was concerned only with research-based links.

To quantify research-based causal inferences, we first identified all research references or formal citations in the texts. We then looked to see if these references were used to support causal inferences about any elements within the five policy design components, as determined previously. To be counted as a link, at least two elements belonging to different, adjacent model components had to be present. The number of links for each location (A–D) in the logic model were counted and compared within and across policy designs (see Figure 2).

To assess accuracy and permit possible revision to these coding rules, we piloted them on the proposed rule for Veteran HIV testing policy that was published in the Federal Register on December 29, 2008 (Department of Veterans Affairs 2008). This early draft of the VA policy invited public

Table 1: *Logic Modeling* of the Five Components of HIV Screening Policy from the Centers for Disease Control and Prevention

(1) Resources/Inputs	*(2) Activities*	*(3) Outputs*	*(4) Outcomes*	*(5) Impact*
Technology: Rapid tests Human: Lack of provider time to conduct risk assessment Patient perceptions and disclosure of risk Interpreters and bilingual staff Financial: Cost of HIV screening often not reimbursed Organizational: Previous CDC recommendations Literature review Lessons learned from CDC-sponsored demonstration projects Setting-specific prevalence data not available Community: Expert meetings	Actions: Ascertain and document all known risk factors Events: Recommendations revised on September 22, 2006 Technology: Use plasma RNA test and HIV antibody test when acute retroviral syndrome is a possibility Services: Routine, voluntary screening as a normal part of medical practice Annual repeat testing of high-risk patients Prevention counseling not required Processes: Screen if population prevalence >0.1% Opt-out screening General medical consent sufficient to encompass HIV testing Written consent not required Separate consent form not recommended Document testing decision in medical record	Service type: Detect by reliable, inexpensive, noninvasive screening test Identify unrecognized health conditions Diagnose infection before symptoms develop Increase awareness of status Service level: Increase HIV screening in all health care settings Make testing more feasible and less costly Service target: Patients aged 13–64 Patients undergoing treatment for TB Patients seeking treatment for STD Patients at high-risk: injection-drug users and their sex partners, persons who exchange sex for money/drugs, sex partners of HIV-infected persons, and MSM or heterosexual persons who themselves or whose sex partners have had >1 partner since most recent test	Short term: Preserve patient's option to decline Definitive mechanism to inform patients of results Negative test results may be conveyed without direct contact Positive test results must be conveyed confidentially in person Confirmatory testing for vaccine trial participants Document results in medical record Foster earlier detection of HIV infection Counsel and link patients who test positive to clinical and prevention services Ensure immediate access to clinical care for persons with newly identified HIV Offer treatment before symptoms develop Reduction in viral load through timely initiation of HAART Implement interventions to reduce transmission risk	Systemic: Produce public health benefit from reduced transmission to sex partners Promote justice Preserve individual rights

continued

Table 1. *Continued*

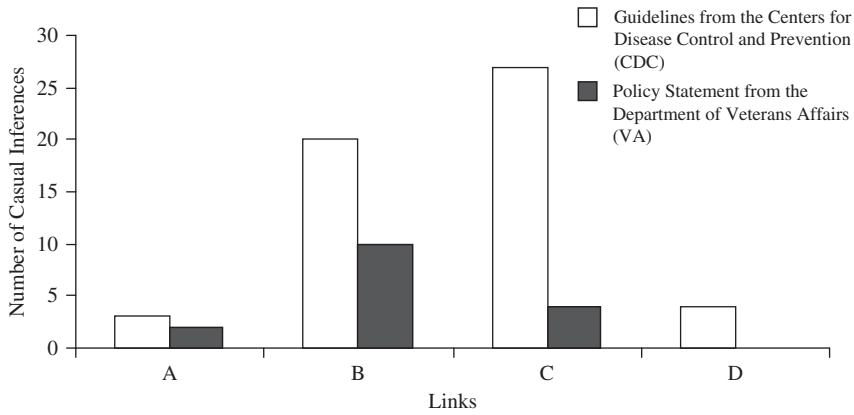
(1) Resources/Inputs	*(2) Activities*	*(3) Outputs*	*(4) Outcomes*	*(5) Impact*
Stakeholder consultation Peer review Invited professional comment Legal: State regulations that limit diagnostic testing for HIV infection	Oral or written information should explain meaning of HIV infection and test results Offer opportunity for questions More frequent diagnostic testing of patients with symptoms Jurisdictions should implement recommendations within current parameters/ consider steps to resolve legal conflicts Products: Multilingual informational materials	Patients considering initiation of new sexual relationship Patient sources of an occupational exposure	Patients decrease high-risk behavior Encourage status disclosure to spouses, current and previous sex partners Health departments can notify/counsel/test partners without disclosing patient identity Report HIV diagnosis/AIDS cases to health department Evaluate patient connection/ continued engagement in care Long term: Ensure that patient–provider relationship is conducive to optimal clinical and preventative care De-stigmatize the testing process Link clinical care with prevention Slower clinical progression Gain years of life with early treatment Reduce mortality Cost-effective where prevalence $\geq$ 0.1% Monitor costs	

*Stated barriers to policy change are given in parantheses.

Table 2: *Logic Modeling* of the Five Components of HIV Screening Policy from the Department of Veterans Affairs

(1) Resources/Inputs	*(2) Activities*	*(3) Outputs*	*(4) Outcomes*	*(5) Impact*
Materials: Written educational materials Technology: HIV website Computerized provider ordering entry system Electronic medical record Electronic reminders to test Community: 30-day comment period Invited public comment Legal: Mental Health and Other Care Improvement Act of 2008 Senate report Proposed rule in Federal Register on December 29, 2008	Actions: Update informed consent requirements Events: Rule effective on August 17 2009 Technology: Utilize technologies for testing prompts Services: Make testing a part of routine care Eliminate pre-test counseling requirement Eliminate posttest counseling requirement Offer annual repeat testing to high-risk patients Counsel high-risk patients Processes: Offer testing to all patients, not just those at high risk Streamline protocols to be less cumbersome for patients and providers Eliminate written consent requirement Document oral consent decision in progress note Provide written educational materials Give thorough and accurate information about HIV and HIV testing Products: Internal policy guidance	Service type: Bring informed consent and testing procedures in line with VA procedures for other routine clinical tests Align with current practice standards Use testing methods recommended by CDC Refer patients at high risk to services Service level: Increase testing rates Increase education of patients about HIV, including methods of transmission Service target: Individuals receiving Veteran Medical Care Benefits	Short term: Refer patients who test positive to state-of-the-art HIV treatment Treat related conditions as soon as possible after diagnosis Reduce high-risk behavior Respect patient privacy Treat veterans with dignity Long term: Improve potential health outcomes of infected patients Prevent complications Protect veterans' rights and interests	Systemic: Advance the country's broader public health goals Protect the rights of patients in other health care systems

Figure 2: Research-Based Causal Inferences in Two Federal HIV Testing Policy Statements



Here, Link A Represents the Number of Causal Inferences That Linked Resources to Activities; B Links Activities to Outputs; C Links Outputs to Outcomes; and D Links Outcomes to Impact. The CDC's Greatest Relative Research Use Emphasized Resulting Outcomes of Policy Change, While the VA's Greatest Relative Research Use Rationalized the Effect of New Policy Activities upon Outputs.

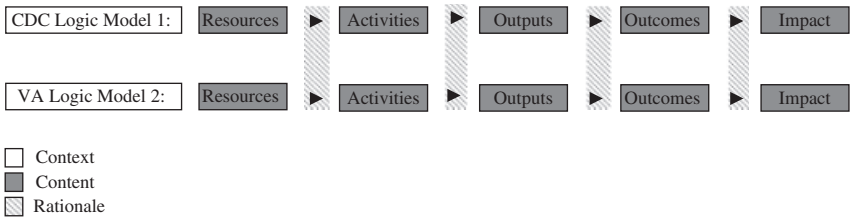
comment prior to the final announcement on July 16, 2009. The result of our pilot evaluation was to assign links occurring between more than two components into all relevant categories; for example, a research-based link between a single output and two or more outcomes would be counted as two pairings for Link C of the model (i.e., both "Output 1 > Outcome 1" and "Output 1 > Outcome 2").

Outcomes of the final comparison were to quantify the number of references given and the number of research-based causal inferences under each link of the model. This was used to determine where agencies were focusing. We incorporated these findings into a conceptual framework connecting policy context, content, and rationale (Figure 3).

RESULTS

CDC and VA HIV testing policy shared design elements under every component of the logic model (see Tables 1 and 2).

Figure 3: Conceptual Framework for Comparative *Logic Modeling* across Two Health Policies



The Structural Components (1) Resources, (2) Activities, (3) Outputs, (4) Outcomes, and (5) Impacts (Gray Boxes) Were Evaluated across the CDC (Top) and VA (Bottom) Logic Models. The Causal Inferences (Black Arrows) Linking Model Components Were Compared with Provide Insights into Research Utilization by Each Agency. Where Supported by Research Citations, These Links Provided a Research-Based Rationale for Comparison across Policies (Diagonal Lines).

Resources

Both agencies invited community input into the design process. Only the CDC mentioned potential barriers to proposed testing activities in the form of human, financial, organizational, and legal resources not be readily available. The VA listed technological resources such as an educational HIV website, electronic medical records, and a computerized provider ordering system that could better enable policy activities.

Activities

Both agencies proposed routine, voluntary HIV testing and removed prior requirements for written consent and pretest counseling. They required oral informed consent and its documentation within the medical record, as well as provision of multilingual educational materials to accompany testing. Both recommended annual repeat testing of patients at high risk for infection. The VA differed from the CDC in that it specifically eliminated a requirement for posttest counseling. The CDC did not require direct personal contact between patient and provider to convey negative test results, but it did recommend counseling referrals for high-risk patients and required efforts that would link patients who test positive to counseling. The CDC proposed additional criteria for screening in populations where HIV prevalence is greater than 1 percent. The CDC noted that where these policy changes were incompatible with existing state laws, steps should be taken to resolve the conflicts.

Outputs

Both agencies anticipated increased screening rates. The CDC recommended that all health care settings implement HIV testing of patients aged 13–64, patients seeking treatment for tuberculosis or sexually transmitted diseases, patients who are starting a new sexual relationship, patients who are sources of occupational exposure, and patients who are thought to be at high risk for infection. The VA made recommendations for Veterans receiving medical care benefits. Here there was a marked difference in policy scope between the disease monitoring agency and VA's health services delivery arm.

Outcomes

Both agencies anticipated a reduction in high-risk behaviors in the short term. Patients who tested positive would be linked to treatment. Long-term goals for both agencies were to reduce HIV-related morbidity and mortality. The CDC foresaw further benefit in removing stigma against HIV testing, ensuring a good patient–provider relationship and achieving cost-effectiveness.

Impact

Both agencies anticipated advancement of public health goals and patient rights.

Research Evidence Application

Both the CDC and the VA used research references to support policy design inferences, but the chosen evidence base differed greatly between these two agencies.

The VA emphasized a need for Veteran-specific research, and, unlike the CDC, cited studies conducted at VA health care facilities and with a primary focus on the U.S. Veteran population. The VA disagreed with public comment that would maintain mandated pre- and posttest counseling because that literature was "drawn from settings outside the VA" (Department of Veterans Affairs 2009).

Whereas the VA was issuing a final rule governing health practices of its own service facilities, the CDC was generating policy recommendations for adoption externally. As such, the CDC required a strong level of transparency to make the guidelines readily adoptable to a variety of U.S. settings.

Consistent with these different policy mandates, we found no overlapping references between HIV testing policy statements, despite similar design components. The VA did not cite the CDC guidelines formally, but stated its

intention to bring procedures in line with current CDC HIV testing recommendations, suggesting that the agency is familiar with the evidence base of CDC policy without the need to reiterate these references.

There were differing quantities of research evidence, with the CDC citing many more references than the VA. It is important to note that this comparison captures only symbolic research references in the final policy statements. The policies may be products of greater instrumental and conceptual research use occurring earlier in the process in order to identify desired evidence-based practices.

We found additional differences between agency policy statements after quantifying and comparing the research-based causal inferences used to link design components. Every link in the CDC logic model was supported with at least one research-based causal inference (Figure 2), but the VA model did not link outcomes to impact. Particular links of the logic model took on greater relative importance according to the agency (Figure 2). In the following, we provided only a few examples of these causal links.

Link A, Resources to Activities

Both agencies cited minimal evidence linking policy resources to activities and focused more on the limitations of resources to produce successful alternatives, such as risk-based screening.

Link B, Activities to Outputs

The VA emphasized this link with the bulk of its research citations. Both agencies believed that an increase in testing rates would result from the elimination of pretest counseling and prior written informed consent requirements.

Link C, Outputs to Outcomes

The majority of research-based causal inferences made by the CDC linked outputs to outcomes. The agency focused on findings of reduced transmission, improved health outcomes, and reasonable cost-effectiveness, which occurred as a result of early HIV diagnosis and treatment. The VA also offered justifications for this model link, stating that research exists which supports an "excellent record of linkage to care" following positive HIV diagnosis.

Link D, Outcomes to Impact

Only the CDC cited research to support an impact to patient rights, justice, and public health over the longer term.

DISCUSSION

Since 2006, the CDC has recommended routine, one-time HIV testing for all U.S. adolescents and adults ages 13–64, in all health care settings except those with undiagnosed HIV prevalence known to be <0.1 percent. Current guidelines support a broad opt-out testing approach under the patient's general consent for medical care, with annual repeat testing of patients at highest behavioral risk (Branson et al. 2006). These guidelines differ from past CDC recommendations that focused screening on those with risk behaviors, required signed, informed consent, and included pre- and posttest behavioral counseling to reduce risk behaviors (Centers for Disease Control and Prevention 2001). The use of *Logic Modeling* as a comparative tool for policy analysis provides a systematic and transparent approach to examine design logic and research use in relating the CDC guidelines to requirements in VA health care facilities that began in 2009.

Prior study has indicated that the policy making and research processes are heavily influenced by context (Rütten et al. 2003; Almeida and Báscolo 2006; Contandriopoulos et al. 2010), and that policy making setting may be a factor in research utilization (Weiss 1978; James and Jorgensen 2009). In this study, we developed a conceptual framework to evaluate context, content, and rationale of the policy process simultaneously (Figure 3).

While we found some differences in CDC and VA policy design and rationale, the agencies' HIV testing policies are largely comparable. The strong overlap of CDC and VA HIV testing policy components in this study suggests that VA policy may be generalizable to other settings considering adoption of, or alignment with, CDC guidelines. The CDC has identified "states, local jurisdictions, or agencies" as the regulatory bodies that oversee HIV screening, in such settings such as "hospital emergency departments, urgent care clinics, inpatient services, substance abuse treatment clinics, public health clinics, community clinics, correctional health care facilities, and primary care settings" (Branson et al. 2006). These settings all stand to benefit from observation of HIV testing policy implementation at the VA.

Content overlap between agencies' policies permitted comparison of research-based causal inferences. Despite similar policy design choices, however, we found no overlapping references between the CDC and VA policy statements. This may indicate the use of other (non-research-based) information types to make policy design choices. But because it reflects symbolic research use by the agencies, this finding provides additional insights into the tactics, formal interests, and design focus of agency policy makers. We found

that the VA cited evidence that was specific to U.S. Veterans in VA health service settings, while the CDC drew from a broader evidence base than the VA, as is reflective of their differing scopes of practice.

We found that the CDC offered evidence in support of all logic model links, while the VA did not justify its claims of a longer-term impact for the proposed policy change. This finding may represent an evidence gap in VA policy rationale, or the agency may be citing supportive evidence as a formality, having already decided to adopt the CDC guidelines. When used symbolically, research utilization in policy statements may be a tactical move designed to create stakeholder buy-in of an earlier, and authoritative, decision making process.

Depending on the agency, particular links of the logic model took on greater relative importance. The majority of the VA's research-based justifications were given to link planned activities to expected outputs, while the CDC placed the most evidentiary emphasis on linking outputs to outcomes, a finding also related to the agencies' different public health missions. The VA is responsible for operations specific to U.S. Veterans' health and has focused on policy throughputs. The CDC is charged with setting broad policy to influence service providers and the public, and it has placed greater emphasis on policy results. These research emphases differ because symbolic research use can achieve different strategic purposes. Policy makers' purpose will determine the need for supportive evidence that justifies the decision. The VA is running a health care system and must directly allocate resources as well as persuade practitioners, patients, and other stakeholders that the new processes and procedures put into place will achieve desired change. Alternatively, the CDC develops and promotes public health policies with the goal of improving care across the U.S. health care system, and it must communicate the wider benefits of adopting new testing guidelines to an array of health care settings.

Our study connects symbolic research utilization to the federal policy making process through HIV testing policy content and rationale. Others have suggested that different types of evidence are useful at different times in the policy process (Bowen and Zwi 2005), and the use of *Logic Modeling* in this study provided a content-based window to compare these contexts, as illustrated by our conceptual framework (Figure 3).

There are several limitations to the approach we have taken here. Policy documents used to reconstruct policy logic are windows into decision making, but they do not capture instrumental or conceptual uses of research. However, it may be valuable to judge the policy at face value—that is, how it is presented in its final form to other agencies, states, service organizations, patients, and

stakeholders who will interpret and apply the policy. Policy statements that provide minimal insight into the intentions, strategies, and rationale of a policy decision are noteworthy for their lack of logic and research-based rationale.

Policy decision making is frequently criticized for a lack of rationality (Buse, Mays, and Walt 2005). This study does not determine if supportive research evidence exists but was not cited, or if policy logic was based on other types of information. Future study might adapt our proposed framework so that logic model links represent multiple information types. Indeed, it is this combination of information sources that is critical to the meaning of evidence-based policy (Bowen and Zwi 2005). Future study also may look explicitly at included and excluded information as it varies across setting with policy design. Others have advocated this approach as context-specific evidence is critical to effective policy making (Bowen and Zwi 2005; James and Jorgensen 2009).

A final limitation of the study is the potential for researcher bias in creating and employing the counts used to quantify research-based links. We reduce such bias by defining our recording units in advance (i.e., the model components and links) and by piloting our coding rules on the proposed VA rule for stability and reproducibility.

CONCLUSION

This case study has compared the 2009 policy change in HIV testing at the VA with existing screening guidelines from the CDC. Through the use of comparative *Logic Modeling* as a tool for policy evaluation, we examined the substance and rationale of policy choices, and the research emphases of these designs in determining an evidence-based practice.

We found considerable overlap in agency policy logic despite dissimilar use of research evidence. The VA largely replicated the CDC HIV testing guidelines, a result which suggests that future evaluation of the VA's policy adoption could be compared to other non-VA health care facilities. In fact, while the VA is unique as one of the world's largest integrated health care systems providing services to U.S. veterans, it is often overlooked as a relevant source of information for other health care organizations. There are often important similarities between the VA and other large health care systems, such as large proportions of older clients in need of chronic disease management, overall size, geographic spread, and level of system integration.

While the CDC and the VA shared policy elements under every component of the logic model, there were no shared research citations used to justify

model links. This difference in research use may be attributable to the distinct missions of the agencies and to the particular focus of the VA on determining evidence-based practice from a Veteran-specific context. This conclusion is further supported by the agencies' different emphases on model links. Recently, *Logic Modeling* was used to evaluate the health policy process in Vietnam across three maternal health case studies (Ha et al. 2010). These research utilization findings suggest that this method may be useful in identifying policy maker interests and intentions in other contexts as well; for example, in comparing policy drafts across legislative bodies, across time and/or settings, such as historical or international comparisons of similar health care services.

The quantification of causal links in *Logic Modeling* is an effective approach to comparing research use in health policy. As we continue to develop new methods for characterizing this research–policy relationship, advancement of our theoretical knowledge through new tools of policy analysis will be critical to the promotion of effective public health policies.

ACKNOWLEDGMNTS

Joint Acknowledgment/Disclosure Statement: This project was funded by the Ruth Freelander Kotlier Graduate Student Enrichment Fund from Boston University, College of Health and Rehabilitation Sciences: Sargent College. The authors wish to thank Ms. Nicole Germino for her editorial and administrative assistance in preparation of the manuscript.

Disclosures: None.

Disclaimers: None.

REFERENCES

Almeida, C., and E. Báscolo. 2006. "Use of Research Results in Policy Decision-Making, Formulation, and Implementation: A Review of the Literature." *Cadernos de Saúde Pública, Rio de Janeiro* 22 (suppl): S7–33.

Bowen, S., and A. B. Zwi. 2005. "Pathways to "Evidence-Informed" Policy and Practice: A Framework for Action." *PLoS Medicine* 2: 600–05.

Branson, B. M., H. H. Handsfield, M. A. Lampe, R. S. Janssen, A. W. Taylor, S. B. Lyss, and J. E. Clark. 2006. "Revised Recommendations for HIV Testing of Adults, Adolescents, and Pregnant Women in Health-Care Settings." *Morbidity and Mortality Weekly Recommendations and Reports* 55: 1–17.

Buse, K., N. Mays, and G. Walt. 2005. *Making Health Policy (Understanding Public Health).* Maidhead, UK: Open University Press.

Centers for Disease Control and Prevention. 2001. "Revised Guidelines for HIV Counseling, Testing, and Referral." *Morbidity and Mortality Weekly Recommendations and Reports* 50: 1–57.

Contandriopoulos, D., M. Lemire, J. L. Denis, and E. Tremblay. 2010. "Knowledge Exchange Processes in Organizations and Policy Arenas: A Narrative Systematic Review of the Literature." *Milbank Quarterly* 88 (4): 444–83.

Department of Veterans Affairs. 2008. "Elimination of Requirements for Prior Signature Consent and Pre-and Post-Test Counseling for HIV Testing. Proposed Rule." *Federal Register* 73 (249): 79428–30.

Department of Veterans Affairs. 2009. "Elimination of Requirement for Prior Signature Consent and Pre-and Post-Test Counseling for HIV Testing. Final Rule." *Federal Register* 74 (135): 34500–03.

Department of Veterans Affairs: Office of Public Health and Environmental Hazards. 2010. "Women Veterans Health Care: Frequently Asked Questions" [accessed on January 1, 2011]. Available at http://www.publichealth.va.gov/womenshealth/faqs.asp

Dobrow, M. J., V. Goel, and R. E. G. Upshur. 2004. "Evidence-Based Health Policy: Context and Utilisation." *Social Science and Medicine* 58 (1): 207–17.

Dwyer, J. 1996. "Applying Program Logic Model in Program Planning and Evaluation." *Public Health and Epidemiology Report Ontario* 7: 38–46.

Ha, B. T. T., A. Green, N. Gerein, and K. Danielsen. 2010. "Health Policy Processes in Vietnam: A Comparison of Three Maternal Health Care Studies." *Health Policy* 98 (2): 178–85.

Hanney, S., M. Gonzalez-Block, M. Buxton, and M. Kogan. 2003. "The Utilisation of Health Research in Policy-Making: Concepts, Examples and Methods of Assessment." *Health Research Policy and Systems* 1 (1): 1–28.

Innovation Network. 2008. *Logic Model Workbook.* Washington, DC: Innovation Network [accessed on January 1, 2011]. Available at www.innonet.org/client_docs/File/logic_model_workbook.pdf.

James, T. E, and P. D. Jorgensen. 2009. "Policy Knowledge, Policy Formulation, and Change: Revisiting a Foundational Question." *Policy Studies Journal* 37 (1): 141–62.

Julian, D. 1997. "The Utilization of the Logic Model as a System Level Planning and Evaluation Device." *Evaluation and Program Planning* 20 (3): 251–57.

Kaplan, S. A., and K. E. Garrett. 2005. "The Use of Logic Models by Community-Based Initiatives." *Evaluation and Program Planning* 28: 167–72.

Kellogg Foundation. 2004. *Logic Model Development Guide: Using Logic Models to Bring Together Planning, Evaluation, and Action.* Battle Creek, MI: W.K. Kellogg Foundation.

Kroneman, M. W., and J. van der Zee. 1997. "Health Policy as a Fuzzy Concept: Methodological Problems Encountered When Evaluating Health Policy Reforms in an International Perspective." *Health Policy* 40 (2): 139–55.

Macaskill, L., J. J. M. Dwyer, C. Uetrecht, C. Dombrow, R. Crompton, B. Wilck, and J. Stone. 2000. "An Evaluability Assessment to Develop a Restaurant Health Promotion Program in Canada." *Health Promotion International* 15 (1): 57–69.

Millar, A., R. S. Simeone, and J. T. Carnevale. 2001. "Logic Models: A Systems Tool for Performance Management." *Evaluation and Program Planning* 24: 73–81.

Moyer, A., H. Verhovsek, and V. L. Wilson. 1997. "Facilitating the Shift to Population-Based Public Health Programs: Innovation through the Use of Framework and Logic Model Tools." *Canadian Journal of Public Health* 88 (2): 95–8.

Owens, D. K., V. Sundaram, L. C. Lazzeroni, L. R. Douglass, P. Tempio, M. Holodniy, G. D. Sanders, V. M. Shadle, V. C. McWhorter, T. Agoncillo, N. Haren, D. Chavis, L. H. Borowsky, E. M. Yano, P. Jensen, M. S. Simberkoff, and S. A. Bozzette. 2007. "HIV Testing of at Risk Patients in a Large Integrated Health Care System." *Journal of General Internal Medicine* 22 (3): 315–20.

Pelz, D. C. 1978. "Some Expanded Perspectives on Use of Social Science in Public Policy." In *Major Social Issues: a Multidisciplinary View*, edited by J. M. Yinger and S. J. Cutler, pp. 347–57. New York: Free Press.

Rich, R. F. 1997. "Measuring Knowledge Utilization: Process and Outcomes." *Knowledge and Policy* 10: 11–24.

Rütten, A., G. Lüschen, T. von Lengerke, T. Abel, L. Kannas, J. A. Rodríguez Diaz, J. Vinck, and J. van der Zee. 2003. "Determinants of Health Policy Impact: A Theoretical Framework for Policy Analysis." *Sozial- und Präventivmedizin* 48 (5): 293–300.

Trostle, J., M. Bronfman, and A. Langer. 1999. "How Do Researchers Influence Decision-Makers? Case Studies of Mexican Policies." *Health Policy and Planning* 14: 103–14.

Weiss, C. 1979. "The Many Meanings of Research Utilization." *Public Administration Review* 39: 429–31.

Weiss, C. H. 1978. "Improving the Linkage between Social Research and Public Policy." In *Knowledge and Policy: The Uncertain Connection*, edited by E. L. Laurence, pp. 23–81. Washington, DC: National Academy of Sciences.

Wholey, J. S. 1987. "Evaluability Assessment: Developing Program Theory." In *Using Program Theory in Evaluation, New Directions for Program Evaluation*, edited by L. Bickman, pp. 77–92. San Francisco: Jossey-Bass.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

BMJ open accessible medical research

Is unhealthy substance use associated with failure to receive cancer screening and flu vaccination? A retrospective cross-sectional study

Karen E Lasser,[1,2] Theresa W Kim,[1,3] Daniel P Alford,[1,3] Howard Cabral,[4] Richard Saitz,[1,3,5] Jeffrey H Samet[1,2,3]

▶ Prepublication history for this paper is available online. To view these files please visit the journal online (http://bmjopen.bmj.com).

Received 18 January 2011
Accepted 2 March 2011

Preliminary findings were presented at the Society of General Internal Medicine conference, 28 April to 1 May 2010, Minneapolis, Minnesota, and the Quality of Behavioral Healthcare Conference, 14 April 2010, Clearwater, Florida.

For numbered affiliations see end of article.

Correspondence to
Dr Karen E Lasser;
karen.lasser@bmc.org

ABSTRACT

Objective: To compare cancer screening and flu vaccination among persons with and without unhealthy substance use.

Design: The authors analysed data from 4804 women eligible for mammograms, 4414 eligible for Papanicolou (Pap) smears, 7008 persons eligible for colorectal cancer (CRC) screening and 7017 persons eligible for flu vaccination. All patients were screened for unhealthy substance use. The main outcome was completion of cancer screening and flu vaccination.

Results: Among the 9995 patients eligible for one or more of the preventive services of interest, 10% screened positive for unhealthy substance use. Compared with women without unhealthy substance use, women with unhealthy substance use received mammograms less frequently (75.4% vs 83.8%; $p<0.0001$), but Pap smears no less frequently (77.9% vs 78.1%). Persons with unhealthy substance use received CRC screening no less frequently (61.7% vs 63.4%), yet received flu vaccination less frequently (44.7% vs 50.4%; $p=0.01$). In multivariable analyses, women with unhealthy substance use were less likely to receive mammograms (adjusted odds ratio 0.68; 95% CI 0.52 to 0.89), and persons with unhealthy substance use were less likely to receive flu vaccination (adjusted odds ratio 0.81; 95% CI 0.67 to 0.97).

Conclusions: Unhealthy substance use is a risk factor for not receiving all appropriate preventive health services.

ARTICLE SUMMARY

Article focus

- Do persons with unhealthy substance use receive breast, cervical and colorectal cancer screening less frequently than persons without unhealthy substance use?
- Do persons with unhealthy substance use receive flu vaccination less frequently than persons without unhealthy substance use?

Key messages

- Women with unhealthy substance use are less likely to receive mammograms than women without unhealthy substance use.
- Persons with unhealthy substance use are less likely to receive flu vaccination than persons without unhealthy substance use.
- Unhealthy substance use is not a risk factor for not receiving cervical or colorectal cancer screening.

Strengths and limitations of this study

- Strengths: the study used validated measures of unhealthy substance use and encompassed a wide range of substance-use severity.
- Limitations: the findings from our sample of an inner-city patient population with health insurance and access to care who receive primary care at an urban safety-net hospital may not be generalisable to other patient populations. The study cannot determine whether unhealthy substance use causes patients not to receive certain services, or whether screening, brief intervention and substance-use treatment led some patients to complete screenings or vaccination. The study did not obtain records of services performed outside Boston Medical Center, and relied on patient self-report of substance use.

INTRODUCTION

Cancer and flu are among the leading causes of mortality in the USA.[1 2] Flu is preventable, in part, through vaccination, and mortality from cervical, breast and colorectal cancer (CRC) can be reduced through routine screening.[3–5] Nevertheless, many eligible US adults do not receive these recommended preventive services,[6] in particular, low-income persons,[7] racial and ethnic minorities,[8–11] the uninsured[12] and the foreign-born.[13] Despite this knowledge, and the implementation of interventions targeting these groups, preventive services are still underused, which has led

some to believe that high-risk 'pockets' of the population may account for gaps in service receipt. Persons with unhealthy substance use (for alcohol, the spectrum that ranges from risky use to dependence; for drugs, the spectrum from any illicit drug use (including prescription drugs) to dependence) may represent one such 'pocket.' Disorganisation, intoxication, comorbid mental illness and low utilisation of primary care among persons with unhealthy substance use[14] might lead to lower use of preventive services.

Prior studies of cancer screening[15–19] and flu vaccination[18 20] suggest that receipt of these services may be low among persons with substance-use disorders (with levels of use that are severe enough to warrant a diagnosis of abuse or dependence). These studies have been limited by their reliance on ICD-9 codes to define substance-use disorders, their exclusion of persons whose substance use is undiagnosed or does not meet criteria for abuse or dependence, and the fact that they have largely been conducted in Veterans Administration (VA) settings, where patients may not be representative of the general population.

We analysed data on unhealthy substance use collected prospectively and systematically by staff whose sole responsibility across a variety of healthcare settings was screening with brief intervention for substance use, and referral to treatment for substance-use disorders. We linked these data to electronic medical record data at eight urban safety-net hospital-based primary care practices to examine preventive service receipt among persons with and without unhealthy substance use. We hypothesised that persons with unhealthy substance use would receive preventive services less frequently than persons without unhealthy substance use.

METHODS

Study setting and sample

Boston Medical Center is an urban safety-net hospital with eight academic primary care practices staffed by 105 primary care practitioners, including both general internists and family practitioners, and staff and resident physicians. The primary care practices predominantly serve a minority and multicultural low-income population. We identified women eligible for breast cancer screening, women eligible for cervical cancer screening, and men and women eligible for CRC screening. Among these groups examined for cancer screening, we also identified individuals eligible for flu vaccination. We linked these four cohorts of patients to unhealthy substance-use screening data that were obtained over a similar time period in the outpatient, inpatient, and emergency department settings.

From 2007 to the present, Boston Medical Center participated in a universal substance-use screening programme supported by the federal government known as the Massachusetts Screening, Brief Intervention, Referral and Treatment (MASBIRT) programme. As part of the programme, trained lay-persons ask the following three questions of all patients in multiple settings to identify unhealthy substance use:

1. In the past 3 months, how often have you had more than four drinks (with alcohol) in a day (for men; women and men 65 years and over were asked about more than three drinks in a day)?
2. In the past 3 months, how often have you used narcotic pain medicines, sedatives (benzodiazepines), or Ritalin/amphetamine without a doctor's prescription or in greater amounts than prescribed?
3. In the past 3 months, how often have you used marijuana, cocaine, heroin or other drugs?

Unhealthy substance use was defined as any response other than 'never' to any of the above questions. In its clinician's guide, the National Institute on Alcohol Abuse and Alcoholism recommends the single-question screen for unhealthy alcohol use (similar to question 1 above).[21] Smith *et al* validated the single-question screen at Boston Medical Center, finding that it is both sensitive and specific for the detection of unhealthy alcohol use.[22] Since brief validated screening questions for illicit drug use or prescription drug misuse in the primary care setting have only recently been published,[23] the MASBIRT programme used screening questions (questions 2 and 3 above) that were derived from the more extensive Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) questionnaire[24] (validated in primary care settings) and the National Household Survey on Drug Use and Health.[25] The MASBIRT programme specifically asked about marijuana, cocaine and heroin, as these are the three most common illicit drugs used in Massachusetts.[26] Instead of past-year use, all screening questions asked about use in the past 3 months to increase the likelihood that a positive test would make logical sense for a clinician to address (current use) and to match the time frame in the ASSIST questionnaire. The ASSIST questionnaire was administered to all patients who reported drug use or risky alcohol use (an affirmative response to question 1 above), providing a measure of current (or risk of developing) substance-related problems. We defined high-risk drug use according to a WHO ASSIST Specific Substance Involvement Score of ≥27, moderate-risk drug use as a score of 4–26 and low-risk drug use as a score of 0–3.[27] Similarly, we defined high-risk alcohol use as a score of ≥27, moderate-risk alcohol use as a score of 11–26 and low-risk alcohol use as a score of 0–10. Patients who screened positive for unhealthy substance use received a single brief counselling intervention and, if indicated, referral for addiction treatment.

We linked clinical information to data on unhealthy substance use among individuals who were screened for unhealthy substance use from 2007 to 2009. We based our eligibility criteria for the cancer-screening measures on modified versions of the corresponding 2007 *Healthcare Effectiveness Data and Information Set* (HEDIS) measures and recommendations of the US Preventive Services Task Force,[3 28–30] and eligibility criteria for flu

vaccination on CDC guidelines.[31] The Boston Medical Center institutional review board approved the study protocol.

Preventive service measures

Using a clinical data warehouse that makes electronic medical records available for research, we identified three groups of patients: (1) female patients aged 21–64 years; (2) female patients aged 42–69 years; and (3) male and female patients age 51–75 years. We chose these age ranges because we sought consistency with the HEDIS measures on cervical- and breast-cancer screening (groups 1 and 2, respectively), and with the United States Preventive Services Task Force recommendations on CRC screening (group 3).[29] Given the questionable value of CRC screening in persons with limited life expectancy,[32] we chose to follow the United States Preventive Services Task Force colorectal cancer recommendations, with age 75 as an upper age limit of screening, rather than age 80, as specified by HEDIS.

We modified the denominator of the cervical cancer screening measure to include any female patient aged 21–64 who had at least one visit to a primary care site at Boston Medical Center in each of the three previous years. We required a minimum of one visit per year to approximate the HEDIS requirement that patients be 'continuously enrolled' in a health plan. The numerator included any patient who received a Papanicolou (Pap) smear in the past 3 years. We excluded women who had undergone a hysterectomy (based on current procedural terminology (CPT) and International Classification of Diseases, version 9 (ICD-9) codes) from both the numerator and the denominator, as Pap smears are rarely indicated in this group.[30]

For the breast-cancer screening measure, we required that female patients aged 42–69 have one visit to a hospital primary care site in each of the two previous years. The numerator included any patient who received a mammogram in the past 2 years. We excluded women who had undergone a bilateral mastectomy or unilateral mastectomy on two separate dates (based on CPT and ICD-9 codes) from both the numerator and the denominator.

For the CRC screening measure, we required that patients aged 51–75 have one visit to a Boston Medical Center primary care site in each of the two previous years. The numerator included any patient who completed home faecal occult blood cards (based on results in the electronic medical record) in the past year, flexible sigmoidoscopy or barium enema in the past 5 years, or colonoscopy in the past 10 years.

We also examined whether patients eligible for cervical, breast and CRC screening who were eligible for flu vaccination were vaccinated. Patients were eligible for flu vaccination as per CDC recommendations during this period if they were aged 65 or older or had one of the following chronic conditions: asthma, chronic obstructive pulmonary disease, congestive heart failure, moderate or severe liver disease, HIV infection, diabetes mellitus or renal insufficiency.

Covariate measures

Guided by Gelberg's adaptation of Andersen's model of health services use,[33] we examined covariates known to affect healthcare utilisation such as gender, race, age, insurance status and language. We defined the burden of medical comorbidity by using the Deyo adaptation of the Charlson Comorbidity Index.[34] Patients were categorised as having significant comorbidity if they had a Charlson–Deyo Score of one or greater. We obtained psychiatric diagnoses from the electronic medical record problem list. In most cases, these diagnoses were made by the patient's primary care provider or by a mental-health specialist. We also examined primary-care utilisation, analysing the number of primary-care visits over the study period.

Statistical methods

Using the SAS computer statistical package, Version 9.1, we performed χ^2 tests to compare differences in preventive-services receipt between persons with and without unhealthy substance use. In exploratory subgroup analyses, we also compared differences in preventive-services receipt between persons with and without unhealthy alcohol use, and with and without any drug use. We used multiple logistic regression to analyse unhealthy substance use as a predictor of receiving each preventive service. Data were missing at random among <5% of all observations. We included all variables in the model based on their a priori clinical significance, and computed adjusted ORs and 95% CIs based on the multiple logistic model. To minimise the potential for collinearity, we examined the variance inflation factor for each covariate. Analyses were conducted using two-sided tests and a significance level of 0.05. We used general estimating equations to account for clustering of patients within clinicians, and clinicians within practices. To detect differences between men and women with unhealthy substance use, we included interaction terms between unhealthy substance use and sex in the multivariable models of CRC screening and flu vaccination.

RESULTS

There were 9995 primary care patients who were eligible for one of the preventive services of interest and had been screened for unhealthy substance use from 2007 to 2009. Table 1 shows the demographic and clinical characteristics of the sample. Patients with unhealthy substance use were slightly younger, and were more likely to be male, English-speaking and of white or black race (vs Hispanic or other race) than were patients without unhealthy substance use. Patients with unhealthy substance use were also less likely to have private insurance and more likely to have Medicaid or Commonwealth Care (a Massachusetts insurance programme for poor and near-poor uninsured adults). Approximately 10% of the sample screened positive for unhealthy substance use. Among these patients, most had unhealthy alcohol use (72.3%), 41.7% had any illicit drug use, and 30.0% had any marijuana

Table 1 Demographic and clinical characteristics of patients engaged in primary care and screened for unhealthy substance use (SU) in Boston, Massachusetts between 2007 and 2009*

Variable	Unhealthy SU n=975 (%)	No unhealthy SU n=9020 (%)	p Value
Mean (SD) age	52.1 (12.3)	54.7 (12.5)	<0.001
Gender, female	52.0	72.6	<0.0001
Language			
English	93.5	73.2	<0.0001
Spanish	4.4	7.6	
Haitian Creole	0.7	9.5	
Other	1.3	9.7	
Race†			
White	21.6	15.9	<0.0001
Black/African−American	63.3	55.3	
Hispanic/Latino	10.8	31.1	
Other	4.3	15.8	
Insurance			
Medicare	29.3	29.8	<0.0001
Health maintenance organization	20.1	26.3	
Medicaid	22.8	17.4	
Free care	5.5	7.7	
Commonwealth care‡	20.0	16.1	
Other	2.3	2.6	
Six or more primary care visits over study period	51.9	53.8	0.25
Significant medical comorbidity§	58.0	54.5	0.04
SU severity			
Unhealthy alcohol use¶	72.3		
Any drug use, past 3 months	41.7		
Marijuana	30.0		
Cocaine	9.0		
Any opioids	7.0		
Drug Involvement Score**			
Low risk	70.4		
Moderate risk	27.2		
High risk	2.5		
Alcohol Involvement Score††			
Low risk	77.4		
Moderate risk	18.5		
High risk	4.1		
Any mental disorder	44.6	35.8	<0.0001
Anxiety	15.4	12.4	0.008
Bipolar disorder	3.9	1.6	<0.0001
Depression	37.4	28.6	<0.0001
Post-traumatic stress disorder	5.9	3.7	0.0006
Panic disorder	1.7	1.5	0.61
Schizophrenia	0.82	1.2	0.28

*Data presented are for unique patients from all four cohorts of patients: (1) women eligible for mammograms (n=4804), (2) women eligible for Papanicolaou tests (n=4414), (3) men and women eligible for colorectal cancer screening (n=7008) and (4) men and women from cohorts 1, 2 and 3 who were eligible for flu vaccination (n=7017).
†Patient race and ethnicity were determined by clinical registration staff.
‡Commonwealth Care, a Massachusetts insurance programme for poor and near-poor uninsured adults.
§Charlson−Deyo Score of ≥1.
¶Defined as more than four drinks with alcohol in 1 day within the past 3 months (for men; more than three drinks with alcohol for women and men over 65 years).
**Risk level based on WHO Alcohol Smoking and Substance Involvement Screening Test Specific Substance Involvement Score. A score of 0−3 is defined as low risk, 4−26 as moderate risk and ≥27 as high risk.
††Risk level based on WHO Alcohol Smoking and Substance Involvement Screening Test Specific Substance Involvement Score. A score of 0−10 is defined as low risk, 11−26 as moderate risk and ≥27 as high risk.

use. Few patients met criteria for high-risk alcohol or drug use (4.1% and 2.5%, respectively). A higher proportion of patients with unhealthy substance use had a mental disorder (p<0.0001) or significant medical comorbidity (p=0.04) relative to patients without unhealthy substance use. Primary care utilisation did not differ among patients with and without unhealthy substance use.

In bivariable analyses, patients with unhealthy substance use were significantly less likely to receive mammograms or flu vaccination than were patients

Table 2 Use of cancer-screening services and flu vaccination according to substance-use characteristics between 2007 and 2009 in Boston, Massachusetts

	Flu vaccination (n=7017), %	Pap smear (n=4414), %	Mammogram (n=4804), %	Colorectal cancer screening (n=7008), %
Substance use				
None	50.4	78.1	83.8	63.4
Unhealthy substance use*	44.7†	77.9	75.4‡	61.7
Unhealthy alcohol use	45.7	79.1	78.2§	61.1
Any drug use	41.7†	75.5	70.0‡	60.8

*Unhealthy alcohol or any drug use.
†Significantly different from persons without unhealthy substance use, χ^2 $p \leq 0.01$.
‡Significantly different from persons without unhealthy substance use, χ^2 $p < 0.0001$.
§Significantly different from persons without unhealthy substance use, χ^2 $p < 0.05$.

without unhealthy substance use (table 2). Women with unhealthy alcohol use were less likely to receive mammograms, while patients with any drug use were less likely to receive flu vaccination or mammograms ($p<0.05$ for all comparisons). Patients with and without unhealthy substance use did not differ in their receipt of colorectal or cervical cancer screening. Among women who were eligible for both a mammogram and a Pap smear, fewer women with unhealthy substance use (56.5%) completed both tests when compared with women without unhealthy substance use (64.5%, p=0.02).

In the multivariable model predicting receipt of mammograms, unhealthy substance use was significantly associated with a lower odds of mammogram receipt (OR 0.69, CI 0.59 to 0.80). Unhealthy substance use was also significantly associated with a lower odds of flu vaccination receipt (OR 0.80, CI 0.66 to 0.97). There were no significant interactions between gender and unhealthy substance use for either CRC screening or flu vaccination. Unhealthy substance use was not an independent predictor of receiving the other preventive services assessed (table 3).

DISCUSSION

Among this sample of patients engaged in primary care, women who screened positive for unhealthy substance use received mammography screening less frequently than women who screened negative. Similarly, men and women who screened positive for unhealthy substance use were less likely to receive flu vaccination than other patients. Not withstanding this identified disparity in the provision of preventive services, delivery of appropriate preventive clinical care in this primary care patient sample was remarkably high, when compared with national estimates.[35] We speculate that persons with unhealthy substance use who are not engaged in primary care at the high thresholds used in these analyses may have substantially lower receipt of preventive services.

Notably, patients with any drug use (which in this study was predominantly marijuana) were also less likely to receive mammography screening and flu vaccination. Because marijuana users are more likely to use tobacco,[36] lower receipt of flu vaccination may have particular clinical significance. Despite large numbers of patients with marijuana use, there are very few studies of

Table 3 Multivariable analyses of the association between unhealthy substance use and receipt of preventive services by primary care patients* between 2007 and 2009 in Boston, Massachusetts

	Flu vaccination OR (95% CI)	Pap smear OR (95% CI)	Mammogram OR (95% CI)	Colorectal cancer screening OR (95% CI)
Unhealthy substance use	0.80 (0.66 to 0.97)	0.95 (0.70 to 1.29)	0.69 (0.59 to 0.80)	0.93 (0.74 to 1.17)
Older age†	1.49 (1.31 to 1.70)	0.30 (0.26 to 0.35)	1.55 (1.26 to 1.90)	0.98 (0.85 to 1.14)
Female	0.74 (0.68 to 0.82)	NA	NA	0.91 (0.80 to 1.04)
Public insurance‡	1.10 (0.98 to 1.24)	0.92 (0.81 to 1.06)	0.86 (0.74 to 0.99)	0.78 (0.66 to 0.93)
Black race	0.79 (0.69 to 0.90)	1.11 (0.98 to 1.26)	1.05 (0.93 to 1.19)	0.94 (0.85 to 1.04)
English-speaking	0.94 (0.77 to 1.14)	0.84 (0.65 to 1.08)	0.75 (0.66 to 0.86)	1.01 (0.84 to 1.22)
Medical comorbidity§	1.54 (1.17 to 2.02)	0.73 (0.57 to 0.93)	0.88 (0.74 to 1.05)	0.98 (0.92 to 1.05)
Psychiatric comorbidity¶	1.20 (1.13 to 1.29)	0.93 (0.74 to 1.18)	0.73 (0.64 to 0.83)	1.04 (0.93 to 1.15)
High primary-care-practice utilisation**	1.89 (1.70 to 2.11)	1.02 (0.78 to 1.33)	1.60 (1.14 to 2.26)	1.59 (1.40 to 1.81)

*The variable unhealthy substance use was included in all models as it is the primary predictor of interest.
†Analyses of flu vaccination receipt compared patients aged 65–75 with those aged 21–64; analyses of Papanicolou (Pap) smear receipt compared patients aged 50–64 with those aged 21–49; analyses of mammogram receipt compared patients aged 50–69 with those aged 40–49; analyses of receipt of colorectal cancer screening compared patients aged 65–75 with those aged 50–64.
‡Defined as Free Care, Medicaid or Commonwealth Care (the new subsidised Massachusetts insurance programme).
§Defined as Charlson–Deyo Score of 1 or greater.
¶Defined as diagnosis of anxiety, bipolar, depression, post-traumatic stress disorder, panic or schizophrenia on medical problem list.
**Defined as at least six primary care visits in the past 2 years for patients eligible for mammograms and flu vaccination, and at least six primary care visits in the past 3 years for patients eligible for Pap smears and colorectal cancer screening.

marijuana and health-services use.[37] Unexpectedly, the proportions of patients with cervical and CRC screening were not lower among persons with unhealthy substance use. It is possible that substance-using women are more likely to have unprotected sex, contract sexually transmitted diseases and then visit a women's health provider who may offer cervical-cancer screening.[38 39] Further, in the medical care system in which this study was performed, a Pap smear can be carried out at the time it is recommended, whereas a mammogram must be scheduled on a different day. This additional requirement to schedule a new appointment on a different day and arrange transportation, and possibly childcare, may explain why women with substance-use disorders had lower odds of receiving mammograms but were no less likely to receive Pap tests. CRC screening is the most complex of the screening services that we examined, potentially requiring advanced scheduling, administration of the preparation and having someone accompany the patient home after the procedure. Thus, there may be other factors more important than substance use affecting its completion. Furthermore, before stating that unhealthy substance use does not appear to be a barrier to completion of this most involved screening test, alternative possibilities merit examination such as disproportionately high numbers with evaluation of gastrointestinal bleeds in this population compensating for fewer with standard screening evaluations.

Prior studies found lower rates of CRC screening among veterans with substance-use disorders.[17 18] The lack of a difference in completion of CRC screening in our study between those with and without unhealthy substance use may be explained by inclusion of the spectrum from mild to severe in that definition, as opposed to limiting substance use to the most severe, those with substance-use disorders. Our finding of a lower frequency of mammogram and flu vaccination receipt and a similar frequency of Pap smear receipt among women with unhealthy substance use is consistent with prior studies.[15 18–20] Our study also showed a lower odds of mammogram receipt among English speakers. It is possible that unmeasured confounders such as low socio-economic status, low health literacy and lower levels of education may account for this finding. Our observation of a lower odds of mammogram receipt among individuals with psychiatric comorbidity is consistent with prior studies,[15] yet contradicts our prior work.[19] In the latter study, primary care and mental-health services were well integrated, which may have accounted for improved preventive screenings among persons with mental illness. It is also possible that individuals with psychiatric comorbidity are more likely to receive preventive services because of their more frequent contact with the health system. Yet, the presence of psychiatric comorbidity can also decrease the likelihood of receiving services if the service requires patient organisation to attend an appointment or to take a preparation. Our finding of a lower odds of Pap tests among women with medical comorbidity is consistent with prior studies.[40]

This study has several limitations. The findings from our sample of an inner-city patient population with health insurance and access to care who receive primary care at an urban safety-net hospital may not be generalisable to other patient populations. Yet, the fact that patients were insured and engaged in primary care helps to isolate the effect of substance use on service receipt. We also cannot determine whether unhealthy substance use causes patients not to receive certain services, or whether screening, brief intervention and substance-use treatment led some patients to complete screenings or vaccination. Further, the periods during which patients were screened for unhealthy substance use and were eligible to receive preventive services overlapped, but some patients may have been screened for unhealthy substance use before or after primary care visits in which prevention was addressed. For example, a patient may have been screened by colonoscopy several years ago, yet was found to have unhealthy substance use more recently. In such cases, it may be difficult to draw conclusions about the association between obtaining a colonoscopy and having substance use. However, the chronic, relapsing and remitting nature of substance use suggests that such use may influence preventive-healthcare utilisation over time.

We did not obtain any records of services performed outside Boston Medical Center. We believe that it is unlikely that patients receiving primary care at Boston Medical Center would have sought and received primary preventive care elsewhere, with the possible exception of the flu vaccine, which is widely available in the community. But even if patients had received services elsewhere, such use would have been associated with non-differential misclassification bias, as we suspect patients with unhealthy substance use are no more likely than other patients to obtain care in other health systems. In multivariable analyses, we observed higher rates of flu vaccination among those with psychiatric comorbidity. It is possible that such patients are less likely than others to seek preventive care outside Boston Medical Center. We relied on patient self-report of substance use. Others have found that self-report of substance use is valid when there are assurances of confidentiality and when validated tools are used.[23] While we used a validated tool, it is possible that some patients under-reported their substance use in the clinical setting. Such under-reporting would have biased our findings to the null.

CONCLUSION

In conclusion, our findings suggest that unhealthy substance use is a barrier to completion of mammography screening and flu vaccination. Future interventions to promote mammography screening might target women with unhealthy substance use, and those to promote flu vaccination might target both men and women with unhealthy substance use. Clinical interventions could embed mammography screening and flu vaccination in

other services delivered to individuals with substance-use problems. Training interventions could enhance skills and systems for healthcare personnel who screen for substance-use disorders to include referrals for preventive health services.

Author affiliations:
[1]Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, Massachusetts, USA
[2]Department of Community Health Sciences, Boston University School of Public Health, Boston, Massachusetts, USA
[3]Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Boston University School of Medicine and Boston Medical Center, Boston, Massachusetts, USA
[4]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA
[5]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA

Acknowledgements The authors thank MM D'Amore, for her assistance in manuscript preparation, L Rosen, for performing data extraction, S Evans, for performing data linkage and data cleaning, and MD Shrayer, for his comments on an earlier draft of the manuscript.

Funding This study was supported by Mentored Research Scholar Grant MRSGT-05-007-01-CPPB from the American Cancer Society (KEL), the Substance Abuse and Mental Health Services Administration (1UTI018311) and the Massachusetts Department of Public Health Bureau of Substance Abuse Services.

Competing interests None.

Contributors KEL conceived and planned the work that led to the article. She wrote the paper and approved the final version. TWK played an important role in interpreting the results, made substantive suggestions for revision and approved the final version. DPA played an important role in interpreting the results, made substantive suggestions for revision and approved the final version. HC played an important role in interpreting the results, made substantive suggestions for revision and approved the final version. RS played an important role in interpreting the results, made substantive suggestions for revision and approved the final version. JHS played an important role in interpreting the results, made substantive suggestions for revision and approved the final version.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Statistical code is available from the corresponding author at Karen.lasser@bmc.org. Consent was not obtained but the presented data are anonymised and risk of identification is low.

REFERENCES

1. US Cancer Statistics Working Group. *United States Cancer Statistics: 1999–2005 Incidence and Mortality Web-based Report*. Atlanta: Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute, 2009.
2. National Cancer Institute. *Common Cancer Types*. Bethesda, Maryland: US National Institutes of Health, 2009.
3. Humphrey LL, Helfand M, Chan BK, *et al*. Breast cancer screening: a summary of the evidence for the US Preventive Services Task Force. *Ann Intern Med* 2002;137:347–60.
4. Winawer SJ, Zauber AG, Ho MN, *et al*. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med* 1993;329:1977–81.
5. Franco EL, Duarte-Franco E, Ferenczy A. Cervical cancer: epidemiology, prevention and the role of human papillomavirus infection. *CMAJ* 2001;164:1017–25.
6. McGlynn EA, Asch SM, Adams J, *et al*. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003;348:2635–45.
7. Armstrong K, Berlin M, Schwartz JS, *et al*. Barriers to influenza immunization in a low-income urban population. *Am J Prev Med* 2001;20:21–5.
8. Centers for Disease Control (CDC). Influenza and pneumococcal vaccination levels among person aged >65 years—United States, 2001. *MMWR Morb Mortal Wkly Rep* 2002;51:1019–24.
9. Centers for Disease Control (CDC). Racial/ethnic disparities in influenza and pneumococcal vaccination levels among persons aged >65 years—United States, 1989–2001. *MMWR Morb Mortal Wkly Rep* 2003;52:958–62.
10. Marin MG, Johanson WG Jr, Salas-Lopez D. Influenza vaccination among minority populations in the United States. *Prev Med* 2002;34:235–41.
11. Egede LE, Zheng D. Racial/ethnic differences in influenza vaccination coverage in high-risk adults. *Am J Public Health* 2003;93:2074–8.
12. Roetzheim RG, Pal N, Tennant C, *et al*. Effects of health insurance and race on early detection of cancer. *J Natl Cancer Inst* 1999;91:1409–15.
13. Mandelblatt JS, Gold K, O'Malley AS, *et al*. Breast and cervix cancer screening among multiethnic women: role of age, health, and source of care. *Prev Med* 1999;28:418–25.
14. Chitwood DD, Sanchez J, Comerford M, *et al*. Primary preventive health care among injection drug users, other sustained drug users, and non-users. *Subst Use Misuse* 2001;36:807–24.
15. Carney CP, Jones LE. The influence of type and severity of mental illness on receipt of screening mammography. *J Gen Intern Med* 2006;21:1097–104.
16. McKnight B, McKnight I, Kerr T, *et al*. Prevalence and correlates of cervical cancer screening among injection drug users. *J Obstet Gynaecol Can* 2006;28:695–9.
17. Aggarwal A, Li N, Lee A, *et al*. Colorectal cancer screening rates in patients with chronic mental illnesses and substance abuse disorders: a retrospective nation-wide data analyses. *Proceedings of the 2009 Annual Meeting of the Society of General Internal Medicine*. Miami, FL: Journal of General Internal Medicine (JGIM), 2009.
18. Druss BG, Rosenheck RA, Desai MM, *et al*. Quality of preventive medical care for patients with mental disorders. *Med Care* 2002;40:129–36.
19. Lasser KE, Zeytinoglu H, Miller E, *et al*. Do women who screen positive for mental disorders in primary care have lower mammography rates? *Gen Hosp Psychiatry* 2003;25:214–16.
20. Merrick EL, Hodgkin D, Garnick DW, *et al*. Unhealthy drinking patterns and receipt of preventive medical services by older adults. *J Gen Intern Med* 2008;23:1741–8.
21. National Institute on Alcohol Abuse and Alcoholism. *Helping Patients Who Drink Too Much: A Clinician's Guide, 2005 Edition*. Bethesda, MD: US Department of Health and Human Services, 2007.
22. Smith PC, Schmidt SM, Allensworth-Davies D, *et al*. Primary care validation of a single-question alcohol screening test. *J Gen Intern Med* 2009;24:783–8.
23. Smith PC, Schmidt SM, Allensworth-Davies D, *et al*. A single-question screening test for drug use in primary care. *Arch Intern Med* 2010;170:1155–60.
24. Humeniuk R, Ali R. *Validation of the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) and Pilot Brief Intervention: A Technical Report of Phase II Findings of the WHO ASSIST Project*. Geneva, Switzerland: World Health Organization, 2006.
25. Office of Applied Studies. *National Survey on Drug Use & Health*. Rockville, MD: US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration (SAMHSA), 2007.
26. Office of Applied Studies. *1993–2003 Treatment Episode Data Set (TEDS) Report of Substance Abuse Treatment Admissions*. Rockville, MD: US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration (SAMHSA), 2007.
27. World Health Organization. *The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) v3.0*. Geneva, Switzerland: WHO Press, World Health Organization, 2010.
28. National Committee for Quality Assurance. *HEDIS: Healthcare Effectiveness Data and Information Set*. Washington, DC: National Committee for Quality Assurance (NCQA), 2007.
29. Screening for colorectal cancer. US Preventive Services Task Force recommendation statement. *Ann Intern Med* 2008;149:627–37.
30. US Preventive Services Task Force. *Screening for Cervical Cancer*. Rockville, MD: Agency for Healthcare Research and Quality (AHRQ), Jan 2003.
31. Centers for Disease Control (CDC). Prevention and control of influenza, recommendations of the advisory committee on immunization practices (ACIP). *MMWR Morb Mortal Wkly Rep* 2003;52:1–34.
32. Walter LC, Lindquist K, Nugent S, *et al*. Impact of age and comorbidity on colorectal cancer screening among older veterans. *Ann Intern Med* 2009;150:465–73.

33. Gelberg L, Andersen RM, Leake BD. The Behavioral Model for Vulnerable Populations: application to medical care use and outcomes for homeless people. *Health Serv Res* 2000;34:1273–302.
34. Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
35. Massachusetts Health Quality Partners. *Quality Insights: Clinical Quality in Primary Care*. Watertown, MA: Massachusetts Health Quality Partners, 2011.
36. Richter KP, Kaur H, Resnicow K, *et al*. Cigarette smoking among marijuana users in the United States. *Subst Abus* 2004;25:35–43.
37. Polen MR, Sidney S, Tekawa IS, *et al*. Health care use by frequent marijuana smokers who do not smoke tobacco. *West J Med* 1993;158:596–601.
38. Rassool G, Villar-Luis M. Reproductive risks of alcohol and illicit drugs: an overview. *J Addic Nurs* 2006;17:211–13.
39. Substance Abuse and Mental Health Services Administration. *The NSDUH Report: Sexually Transmitted Diseases and Substance Use*. Rockville, MD: Office of Applied Studies, 2007.
40. Kiefe CI, Funkhouser E, Fouad MN, *et al*. Chronic disease as a barrier to breast and cervical cancer screening. *J Gen Intern Med* 1998;13:357–65.

HEALTH CARE REFORM

Colorectal Cancer Screening Among Ethnically Diverse, Low-Income Patients

A Randomized Controlled Trial

Karen E. Lasser, MD, MPH; Jennifer Murillo; Sandra Lisboa, BA; A. Naomie Casimir, BA; Lisa Valley-Shah, RN, BSN, MM, CGRN; Karen M. Emmons, PhD; Robert H. Fletcher, MD; John Z. Ayanian, MD, MPP

Background: Patient navigators may increase colorectal cancer (CRC) screening rates among adults in underserved communities, but prior randomized trials have been small or conducted at single sites and have not included substantial numbers of Haitian Creole–speaking or Portuguese-speaking patients.

Methods: We identified 465 primary care patients from 4 community health centers and 2 public hospital–based clinics who were not up-to-date with CRC screening and spoke English, Haitian Creole, Portuguese, or Spanish as their primary language. We enrolled participants from September 1, 2008, through March 31, 2009, and followed them up for 1 year after enrollment. We randomly allocated patients to receive a patient navigation–based intervention or usual care. Intervention patients received an introductory letter from their primary care provider with educational material, followed by telephone calls from a language-concordant navigator. The navigators offered patients the option of being screened by fecal occult blood testing or colonoscopy. The primary outcome was completion of any CRC screening within 1 year. Secondary outcomes included the proportions of patients screened by colonoscopy who had adenomas or cancer detected.

Results: During a 1-year period, intervention patients were more likely to undergo CRC screening than control patients (33.6% vs 20.0%; $P<.001$), to be screened by colonoscopy (26.4% vs 13.0%; $P<.001$), and to have adenomas detected (8.1% vs 3.9%; $P=.06$). In prespecified subgroup analyses, the navigator intervention was particularly beneficial for patients whose primary language was other than English (39.8% vs 18.6%; $P<.001$) and black patients (39.7% vs 16.7%; $P=.004$).

Conclusions: Patient navigation increased completion of CRC screening among ethnically diverse patients. Targeting patient navigation to black and non–English-speaking patients may be a useful approach to reducing disparities in CRC screening.

Trial Registration: clinicaltrials.gov Identifier: NCT01141114

Arch Intern Med. 2011;171(10):906-912

COLORECTAL CANCER (CRC) is the second leading cause of cancer death in the United States and is preventable through screening.[1-3] Nevertheless, approximately 40% of eligible adults in the United States and more foreign-born US residents[4] are overdue for CRC screening.[5] Patients at greatest risk for not being screened include racial minorities,[6] patients with Medicaid or no health insurance coverage,[7] those who are foreign born,[8] and patients with low socioeconomic status.[9] Factors that may contribute to low screening rates among the urban poor with health insurance coverage and access to health care include lack of trust in physicians, an absence of symptoms, fatalistic views regarding cancer,[10] and the lack of a recommendation from a physician for screening.[10]

See Invited Commentary at end of article

Patient navigation is a way to address these barriers to screening. Patient navigators are laypersons from the community who guide patients through the health care system so that they receive appropriate services.[11] The navigators perform a wide range of advocacy and coordination activities, such as assisting patients in obtaining health insurance coverage or transportation to appointments.[12] Using flexible problem solving (rather than provision of a discrete set of services), patient navigators educate patients regarding the disease in question and address the needs of the individual patient. Finally, patient navigators provide social and emotional support to patients.

Several nonrandomized studies,[13-17] including our own,[13] have shown that patient navigation can increase rates of CRC

Author Affiliations are listed at the end of this article.

screening among urban, racial minority patients. One prior randomized study[18] of patient navigation for CRC screening was conducted at a single health center serving mostly patients who speak English or Spanish as their primary language. To build on this limited research, we conducted a randomized controlled trial of patient navigation that included immigrants from the Azores, Brazil, Haiti, and Portugal receiving care at 4 different health centers and 2 public hospital–based clinics in the safety-net health care system (ie, a health care system that provides a significant level of care to low-income, uninsured, and vulnerable populations).

METHODS

STUDY DESIGN

We performed a randomized controlled trial of patient navigation to promote CRC screening. We included a sample of patients aged 52 to 74 years who had not completed CRC screening according to US Preventive Services Task Force guidelines[19] (defined as colonoscopy in the past 10 years, sigmoidoscopy or double-contrast barium enema in the past 5 years, or fecal occult blood testing [FOBT] in the past year) and who spoke English, Haitian Creole, Portuguese, or Spanish as their primary language. We randomly assigned patients to receive a maximum of 6 hours of patient navigation in 6 months or usual care. The intervention encouraged FOBT or colonoscopy, the most frequently used screening tests at the study sites. The Cambridge Health Alliance Institutional Review Board approved the study protocol. The board provided a waiver of informed consent because the study was promoting an established screening standard.

SETTING AND PARTICIPANTS

We conducted the study at Cambridge Health Alliance, a primary care practice–based research network[20] composed of 15 community health centers. The health centers serve a multicultural, low-income population in Cambridge, Somerville, and Everett, Massachusetts. After pilot testing our intervention at 1 health center, we selected 6 primary care sites (4 health centers and 2 hospital-based clinics) that were not part of a Massachusetts Department of Public Health Patient Navigation Program to participate in the study. At the time of the study, all study sites used a common electronic health record (Epic Systems Corporation, Verona, Wisconsin) that supports computerized ordering of laboratory tests and referrals. Gastroenterologists perform colonoscopies at 1 of 3 hospital-based endoscopy centers.

We used the electronic clinical data system to identify patients at the primary care sites who were not up-to-date with CRC screening. We recruited patients aged 52 to 74 years who had had 1 visit to a primary care provider (PCP), ie, a physician or nurse practitioner, in each of the 2 previous years at 1 of the study sites. We limited the sample to patients who identified English, Haitian Creole, Portuguese, or Spanish, the most frequently spoken languages at Cambridge Health Alliance, as their primary language. Using explicitly predefined exclusion criteria, 1 of the study investigators (K.E.L.) performed manual medical record reviews, excluding patients if they had acute illness, an end-stage medical disease,[21] severe psychiatric conditions (such as psychosis, bipolar disorder, paranoia, or schizophrenia), active substance abuse, or cognitive impairment. The **Figure** shows the number of patients accrued, randomized, and assessed for the primary outcome. Of the 823 potentially eligible patients, 465 (56.5%) met study entry criteria.

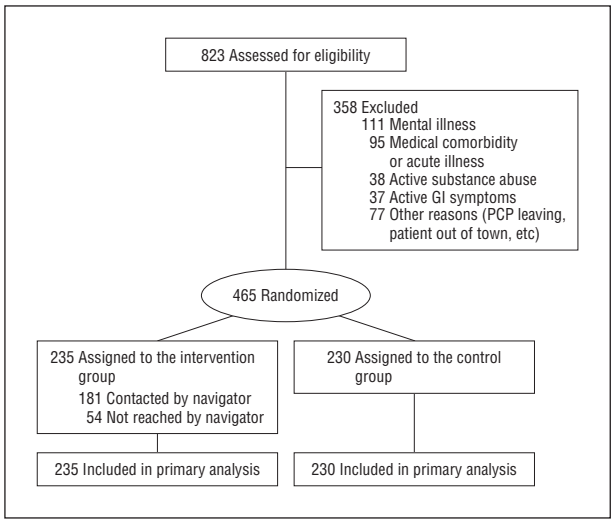


Figure. Consolidated Standards of Reporting Trials study flow diagram. GI indicates gastrointestinal; PCP, primary care provider (ie, a physician or nurse practitioner).

INTERVENTION

We randomly assigned patients to usual care or a maximum of 6 hours of patient navigation during a 6-month period. Some patients required little or no navigation. For example, those who had a previously scheduled colonoscopy appointment and reported that they understood the preparation instructions did not require navigation. For other patients, the maximum of 6 hours was spent contacting patients, educating them regarding CRC and CRC screening tests (FOBT and colonoscopy), motivating them to get screened, helping them decide which test to undergo, helping them obtain health insurance coverage, educating them regarding the correct way to complete FOBT cards, helping them make colonoscopy appointments and finding someone to accompany them home after the procedure, educating them regarding the required bowel preparation, and meeting them on the day of their colonoscopy.

We randomized individual patients, stratified by health center and primary language, using a computer-generated random number table. We enrolled participants from September 1, 2008, through March 31, 2009, and followed them up for 1 year after enrollment. In the intervention group, we sent letters by first-class mail, signed by the PCP of each patient, notifying patients that they were overdue for CRC screening and that a patient navigator would call them. The mailing also included a CRC screening brochure designed by the Harvard Center for Cancer Prevention and the Massachusetts Colorectal Cancer Working Group (*Take Control: Get Tested for Colorectal Cancer*).[22] The brochure, written at a sixth-grade reading level, offered information regarding reasons for screening and different screening modalities. We sent brochures to patients in English, French (for Haitian patients because the brochure was not available in Haitian Creole), Portuguese, or Spanish.

Intervention patients were also eligible to receive navigation from 3 trained navigators who were fluent in English and Spanish, Portuguese, or Haitian Creole. The navigators were based centrally in the Departments of Medicine and Community Affairs. The navigator who worked with Spanish-speaking patients was from Nicaragua, had completed some college-level education, had extensive experience performing community health outreach, and was also a certified nurse assistant. The Portuguese-speaking navigator had a bachelor's degree in clinical psychol-

ogy from Brazil and was an experienced community health worker. The Haitian Creole–speaking navigator was also an experienced community health worker who had completed some college-level education. The navigators were aged 48, 43, and 25 years, respectively. We did not encounter problems matching female navigators with male patients.

The navigators attended a 2-day training program October 4 and 5, 2007, and received additional training on August 22, 2008. The training program included lectures and role-play scenarios concerning the principles of motivational interviewing[23]; CRC and how patients can be screened for it; logistics (eg, how-tos, pros, and cons) of FOBT cards and colonoscopy; CRC prevention (including removal of adenomas); use of open-ended questions, reflective listening, and summarizing; assessment of the readiness of a patient for screening; and approaches for patients who refuse screening (precontemplation), are willing to consider it (contemplation), or are ready to act (action).[23] We based the intervention on a "stages of change" model because other cancer prevention studies[24] have successfully used this model.

During study implementation, the project manager (who also attended the training sessions) audited at least 5 patient calls by each navigator for adherence to a calling script and for motivational interviewing techniques. The patient navigators and the project manager also met on a weekly basis to discuss challenges arising during the outreach calls and to review the use of motivational interviewing techniques.

After randomization, the navigators contacted intervention patients during a 6-month period using a staged rollout procedure by site. Because of financial difficulties, 1 health center closed during the study period (with some patients transferring care to another study health center), and 2 study health centers each merged with other sites that were not part of the study. The navigators continued the intervention with patients at affected sites. Some patients in 2 of the study health centers received mailed outreach material from health center staff regarding CRC screening in early 2009, when planned care outreach became the community standard of care.

During a 3-week period, the patient navigators made as many as 11 attempts to call each patient on different days and times (including evenings and weekends). The navigators also left at least 2 messages for the patient on voice mail or with a family member. If the navigators were unable to contact patients initially, they sent a follow-up letter and made periodic attempts to contact patients during the remainder of the intervention period.

During the initial telephone contact, the navigators educated patients regarding CRC screening, explored barriers to screening, and addressed patients' barriers and their stage of change. For those who were contemplating screening, navigators discussed the screening options of colonoscopy and FOBT cards and the advantages and disadvantages of each test. The navigators presented the screening options in a neutral fashion and did not emphasize the superiority of any test compared with another. The navigators used flexible problem-solving techniques. For example, an elderly patient resisted the idea of colonoscopy because she was concerned that the preparation would make her weak and at risk for a fall. The navigator elicited this concern and suggested the alternative option of FOBT cards.

For patients who chose to complete FOBT cards, the navigator reviewed instructions and mailed FOBT cards and illustrated instructions via first-class mail. If a patient did not return the FOBT cards within 4 weeks, the navigator called the patient to provide support and address barriers to completion. For patients who opted for colonoscopy, the navigators described the test in detail, and the lead navigator contacted the PCP of the patient via the electronic medical record to arrange a colonoscopy referral. Patients with few comorbid medical conditions were referred directly for colonoscopy (also known as open-access colonoscopy); the patient did not require an initial visit with a gastroenterologist but met him or her on the day of the procedure. Patients with conditions such as sleep apnea or a history of problems with anesthesia were referred to a gastroenterologist to discuss colonoscopy. For patients referred directly to colonoscopy, a registered nurse or the lead navigator educated them regarding the procedure and the required bowel preparation, mailed instructions for bowel preparation, and scheduled the procedure.

The navigator worked with the nurse to schedule colonoscopies because the nurse oversaw the colonoscopy schedule. Gastroenterology staff placed reminder calls to all patients 1 day before their colonoscopy. If the navigators were unable to help patients to identify someone to escort them home after the colonoscopy, the navigators advised them to complete FOBT cards instead. For medicolegal reasons, the navigators did not accompany patients home. In some instances, the navigators met patients in the colonoscopy suite to offer emotional support. Some patients were having the first colonoscopy of their life and were fearful about the procedure.

For patients who were not contemplating performing any CRC screening test, the navigators assessed and addressed barriers. For example, for patients who did not have health insurance coverage, the navigators referred them to insurance counselors and worked closely with the insurance department to ensure that eligible patients received coverage. For other patients, the navigators would establish rapport by learning more about the lives of the patients (eg, whether patients were caring for grandchildren or an elderly parent). The navigators motivated some patients to undergo screening by pointing out that by taking care of their own health, patients would be able to continue caring for their family members.

OUTCOME MEASURES

Using a predefined algorithm, 1 study investigator (K.E.L.) performed medical record reviews masked to the intervention assignments to determine CRC screening rates within 1 year of the start of the intervention for all randomized patients. Medical records were reviewed twice to minimize error. We chose to analyze the data at 1 year because the average wait for a screening colonoscopy at the time of the study was approximately 4 months and we assumed that patients would have had sufficient time to complete their colonoscopy during the 1-year period. The primary outcome was completion of CRC screening (colonoscopy, FOBT, flexible sigmoidoscopy, or double-contrast barium enema) within 1 year. Although the navigators focused on helping patients choose between FOBT cards and colonoscopy, PCPs may have chosen to use other accepted screening modalities, namely, flexible sigmoidoscopy or double-contrast barium enema. Secondary outcomes included the proportion of patients screened by colonoscopy and the proportion in whom the screening detected adenomas or cancer. Another secondary outcome was the proportion of patients with high-risk lesions (classified as a dichotomous yes/no variable), defined as 3 or more adenomas of any type, adenomas 1 cm or larger, or adenomas with any villous features.

STATISTICAL ANALYSIS

We calculated that a sample size of at least 197 patients in each group would be required to show a minimum clinically important improvement in CRC screening of 10.0% (10.0% vs 20.0%), with a power of 80.0% and a 2-sided significance level of .05. We conducted all analyses on an intention-to-treat basis. Using χ^2 and Fisher exact tests for dichotomous variables, we performed prespecified subgroup analyses according to the

primary language, age, race, and health insurance coverage status of the patient because these factors are known to affect CRC screening rates.[25] All analyses were performed using SAS statistical software, version 9.1 (SAS Institute Inc, Cary, North Carolina), and we report 2-tailed *P* values or 95% confidence intervals for all comparisons.

RESULTS

PATIENT CHARACTERISTICS

The navigators contacted 181 of the 235 intervention patients (77.0%). The navigators made a median of 8.5 telephone calls (interquartile range, 7.0) to these patients, spending an average of 107 minutes (range, 4-335 minutes) on the telephone with each patient during the 6-month study period. The patient navigators were unable to contact the remaining 54 patients (22.9%) after a median of 10 attempted telephone calls (interquartile range, 7.0), including 34 patients whose telephone service had been intermittently disconnected and 8 patients whose initial outreach letters were returned to sender. Among patients who were and were not contacted, no statistically significant differences were observed in age, sex, race, primary language, or health insurance coverage status.

Table 1 lists the baseline demographic characteristics of the intervention and usual care groups; randomization resulted in groups with similar baseline characteristics. The mean age of the patients was 61.3 years. Most patients were women because women are more likely than men to seek primary care. A substantial proportion of patients were racial minorities and did not have private health insurance coverage. Almost half of patients in both groups spoke English as their primary language (48.2%), with the remainder speaking Portuguese (20.0%), Spanish (13.8%), or Haitian Creole (18.1%).

Table 1. Baseline Patient Characteristics[a]

Characteristic	Intervention, % (n=235)	Control, % (n=230)
Age, mean (SD), y	61.1 (6.0)	61.6 (6.2)
Female	60.4	62.6
Race		
White	47.7	47.4
Black	26.8	28.7
Other	16.6	18.7
Unknown	8.9	5.2
Primary language		
English	47.7	48.7
Portuguese	20.4	19.6
Spanish	14.0	13.5
Haitian Creole	17.9	18.3
Health insurance coverage type		
Private	32.3	33.5
Medicare	21.7	20.0
Medicaid	19.1	20.9
Commonwealth Care[b]	12.8	7.4
Health Safety Net[c]	10.2	11.7
Uninsured	2.5	5.2
Other	1.3	1.3

[a] These characteristics were not statistically significantly different between groups based on χ^2 tests for dichotomous variables and *t* tests for continuous variables. Data are presented as percentages unless otherwise indicated. Percentages may not total 100 because of rounding.

[b] Low-cost or no-cost health insurance coverage program for Massachusetts residents implemented in 2006 as part of the Massachusetts Health Reform Act.

[c] A health insurance coverage program for Massachusetts residents who are not eligible for private coverage or cannot afford it. It replaced the Uncompensated Care Pool (also called Free Care) on October 1, 2007.

OUTCOME MEASURES

A total of 33.6% of intervention patients had been screened by 1 year after study entry vs 20.0% of control patients (**Table 2**; $P<.001$). Intervention patients whom the navigators were able to contact were significantly more likely to be screened than those whom the navigators were unable to contact (39.8% vs 18.6%; $P<.001$). A larger proportion of intervention patients were screened by colonoscopy relative to controls (26.4% vs 13.0%; $P<.001$); similar numbers of intervention and control patients completed FOBT cards (7.2% vs 6.5%; $P=.76$). One control patient had a positive FOBT result that was not followed up with a colonoscopy; the remaining FOBT results were negative. Stratified analyses demonstrated that patient navigation was particularly beneficial for individuals whose primary language was other than English and for patients older than 60 years (Table 2). The intervention was more effective in white and black individuals relative to those of other or unknown race, and a somewhat smaller effect was observed in those with nonprivate vs private health insurance coverage. Because of the small numbers in some race categories, differences between individual race categories should be interpreted with caution.

Intervention patients were more likely to have adenomas detected than were controls (8.1% vs 3.9%; $P=.06$). High-risk adenomas were more frequently detected among intervention patients than among controls (2.5% vs 0.4%; $P=.06$). No colorectal adenocarcinomas were detected in either group.

COMMENT

Patient navigation increased CRC screening rates substantially among racially and linguistically diverse patients served by urban community health centers and public hospital–based clinics. Our study confirms the findings of smaller trials[15,17,20,26,27] and of a recent study[28] of culturally tailored telephone counseling by community health advisers and demonstrates that patient navigation is also effective among patients who speak Haitian Creole or Portuguese as their primary language. Because the Cambridge Health Alliance was undergoing serious financial problems, which led to health center closures and the departure of PCPs, we speculate that our intervention might have had an even stronger effect in a more stable health care system.

The fact that our intervention was effective may reflect the inclusion of several evidence-based components recommended by the Task Force on Community Preventive Services of the Centers for Disease Control and Prevention.[29] These components include client reminders through outreach letters, one-on-one education by the navigators, reduction of structural barriers

Table 2. Receipt of Colorectal Cancer Screening by Intervention Status

Characteristic	No. of Patients (N=465)	Patients Screened, %		*P* Value
		Intervention (n=235)	Control (n=230)	
All patients	465	33.6	20.0	<.001
Age group, y				
50-60	237	27.5	22.2	.35
61-75	228	40.0	17.7	<.001
Sex				
Female	286	33.1	21.5	.03
Male	179	34.4	17.4	.01
Race				
White	221	33.9	16.5	.003
Black	129	39.7	16.7	.004
Other	82	28.2	30.2	.84
Unknown	33	23.8	33.3	.55
Primary language				
English	224	26.8	21.4	.35
Other than English	241	39.8	18.6	<.001
Health insurance coverage type				
Private	153	43.4	22.1	.005
Nonprivate	312	28.9	18.9	.04

related to linguistic and cultural factors, and the inclusion of a mailed educational pamphlet. Although some of the patients in our study may not have been able to read the educational brochure, patients often showed the brochure to their PCP or to a more educated family member. Our intervention also included tailoring to the preferences of patients[30] and informed decision making,[31] elements that may have increased the willingness of patients to be screened. Other possibly beneficial features of the intervention included the lead navigators working closely with the nurse in the gastroenterology center who scheduled colonoscopies; navigators' communicating with PCPs via the electronic medical record, enabling the PCP to place orders for a screening test as soon as a patient chose one; and navigators' working evenings and weekends for greater flexibility in contacting patients. At the time of the intervention, the overall screening rate at the Cambridge Health Alliance was higher than the national average (61% in calendar year 2009) because it reflects patients who are engaged in primary care and who mostly have health insurance coverage. Despite the high baseline screening rate, our intervention was effective.

Strengths of our study are its inclusion of a racially diverse sample of patients from multiple health centers and public hospital–based clinics and its real-world setting. Our study has several limitations. Patients were from 1 geographic area, and some may have obtained CRC screening outside the Cambridge Health Alliance. The navigators were unable to contact 23.0% of intervention patients, which is not surprising given the mobile nature of urban immigrants with low socioeconomic status.[13] In the future, contact rates might be improved by using technology such as text messaging. Patients at some of the intervention health centers began to receive mailed outreach materials regarding CRC screening in early 2009; this may have diminished our intervention effect. We did not have access to data regarding country of birth, citizenship status, length of time in the United States, or level of acculturation for participants. We were also unable to determine which individual components of the intervention were most effective. Prior studies[32,33] have shown that letters have only a limited effect, suggesting that the navigation efforts may have largely accounted for the effectiveness of the intervention. Finally, we excluded patients with active substance use and mental illness documented on the problem list (18.1% of eligible patients); thus, our results may not be generalizable to these patients.

Patient navigation appears to be effective in increasing receipt of colorectal and cervical cancer screening.[34] Yet, clearly, hiring individual navigators for each type of screening would not be feasible. In Massachusetts, as part of a Department of Public Health Program, patient navigators now handle a variety of cancer screenings (breast, cervical, and colorectal) and also provide navigation to link patients to smoking cessation services. Future studies will need to determine whether such dissemination of patient navigation activities, in which an individual patient may receive far less than 1 to 2 hours of patient navigation regarding CRC screening, for example, still results in increased cancer screening. Finally, a need exists to examine the relative efficacy and cost-effectiveness of navigation compared with other viable alternatives (or complements), such as computerized tailored interventions offered in simple and easy-to-understand formats and in the primary language of the user.

As primary care practices are redesigned as medical homes,[35] nonphysician members of the health care team will increasingly take on tasks previously performed by PCPs (eg, counseling and connecting to services). Although the medical home concept is being adopted throughout the United States, its promise requires effective interventions that can be implemented within this new model. Patient navigation is 1 potential intervention that can be integrated into the medical home model. Again, future studies will need to explore whether patient navigation is effective within the context of the medical home.

In conclusion, our findings suggest that patient navigation may represent a powerful tool for increasing CRC screening rates among racially diverse patients. Focusing patient navigation on populations of patients who are black and whose primary language is other than English may be a particularly effective approach to reducing CRC screening disparities for these patients. Future research should assess how health care systems can sustain this benefit when patient navigation is implemented as a routine component of primary care.

Accepted for Publication: November 1, 2010.
Author Affiliations: Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine and Department of Community Health Sciences, Boston University School of Public Health (Dr Lasser), Department of Biomedical Sciences, Boston University School of Medicine (Ms Casimir), Dana Farber Cancer Institute/Harvard School of Public Health (Dr Emmons), Department of Population Medicine/Harvard Medical School and Harvard Pilgrim Health Care (Dr Fletcher), and Division of General Medicine, Brigham and Women's Hospital and Department of Health Care Policy, Harvard Medical School (Dr Ayanian), Boston, Massachusetts; and Departments of Medicine (Ms Murillo) and Community Affairs (Ms Lisboa) and Division of Gastroenterology (Ms Valley-Shah), Cambridge Health Alliance and Harvard Medical School, Cambridge, Massachusetts.
Correspondence: Karen E. Lasser, MD, MPH, 801 Massachusetts Ave, Ste 2091, Boston, MA 02118 (Karen.lasser@bmc.org).
Author Contributions: *Study concept and design:* Lasser, Emmons, Fletcher, and Ayanian. *Acquisition of data:* Lasser, Murillo, Lisboa, Casimir, and Valley-Shah. *Analysis and interpretation of data:* Lasser, Emmons, Fletcher, and Ayanian. *Drafting of the manuscript:* Lasser. *Critical revision of the manuscript for important intellectual content:* Murillo, Lisboa, Casimir, Valley-Shah, Emmons, Fletcher, and Ayanian. *Statistical analysis:* Lasser and Ayanian. *Obtained funding:* Lasser. *Administrative, technical, and material support:* Murillo, Lisboa, Casimir, and Valley-Shah. *Study supervision:* Murillo, Emmons, Ayanian, and Fletcher.
Financial Disclosure: None reported.
Funding/Support: This study was supported by Mentored Research Scholar Grant MRSGT-05-007-01-CPPB from the American Cancer Society.
Role of the Sponsors: This study was supported by the American Cancer Society. The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, and approval of the manuscript.
Previous Presentations: This work was presented at the Society of General Internal Medicine Annual Meeting; April 30, 2010; Minneapolis, Minnesota; and at the Academy Health Annual Research Meeting; June 29, 2010; Boston, Massachusetts.
Additional Contributions: Meredith Manze D'Amore, MPH, provided assistance in manuscript preparation; Maxim D. Shrayer, PhD, provided comments on an early draft of the manuscript; neither individual was compensated. The Community Outreach and Enrollment Department at Cambridge Health Alliance provided assistance in helping patients obtain health insurance coverage.

REFERENCES

1. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood: Minnesota Colon Cancer Control Study [published correction appears in *N Engl J Med*. 1993;329(9):672]. *N Engl J Med*. 1993;328(19):1365-1371.
2. Selby JV, Friedman GD, Quesenberry CP Jr, Weiss NS. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med*. 1992;326(10):653-657.
3. Centers for Disease Control and Prevention (CDC). Vital signs: colorectal cancer screening among adults aged 50-75 years—United States, 2008. *MMWR Morb Mortal Wkly Rep*. 2010;59:1-5.
4. Shih YC, Elting LS, Levin B. Disparities in colorectal screening between US-born and foreign-born populations: evidence from the 2000 National Health Interview Survey. *J Cancer Educ*. 2008;23(1):18-25.
5. Steinwachs D, Allen JD, Barlow WE, et al. National Institutes of Health State-of-the-Science Conference statement: enhancing use and quality of colorectal cancer screening. *Ann Intern Med*. 2010;152(10):663-667.
6. Walsh JM, Kaplan CP, Nguyen B, Gildengorin G, McPhee SJ, Pérez-Stable EJ. Barriers to colorectal cancer screening in Latino and Vietnamese Americans: compared with non-Latino white Americans. *J Gen Intern Med*. 2004;19(2):156-166.
7. Walsh JME, Posner SF, Perez-Stable EJ. Colon cancer screening in the ambulatory setting. *Prev Med*. 2002;35(3):209-218.
8. Goel MS, Wee CC, McCarthy EP, Davis RB, Ngo-Metzger Q, Phillips RS. Racial and ethnic disparities in cancer screening: the importance of foreign birth as a barrier to care. *J Gen Intern Med*. 2003;18(12):1028-1035.
9. Weller D, Coleman D, Robertson R, et al. The UK colorectal cancer screening pilot: results of the second round of screening in England. *Br J Cancer*. 2007;97(12):1601-1605.
10. Lasser KE, Ayanian JZ, Fletcher RH, Good M-JD. Barriers to colorectal cancer screening in community health centers. *BMC Fam Pract*. 2008;9:15-22.
11. Dohan D, Schrag D. Using navigators to improve care of underserved patients: current practices and approaches. *Cancer*. 2005;104(4):848-855.
12. Freeman HP. Patient navigation: a community based strategy to reduce cancer disparities. *J Urban Health*. 2006;83(2):139-141.
13. Lasser KE, Murillo J, Medlin E, et al. A multilevel intervention to promote colorectal cancer screening among community health center patients: results of a pilot study. *BMC Fam Pract*. 2009;10:37-43.
14. Chen LA, Santos S, Jandorf L, et al. A program to enhance completion of screening colonoscopy among urban minorities. *Clin Gastroenterol Hepatol*. 2008;6(4):443-450.
15. Nash D, Azeez S, Vlahov D, Schori M. Evaluation of an intervention to increase screening colonoscopy in an urban public hospital setting. *J Urban Health*. 2006;83(2):231-243.
16. Myers RE, Hyslop T, Sifri R, et al. Tailored navigation in colorectal cancer screening. *Med Care*. 2008;46(9)(suppl 1):S123-S131.
17. Ma GX, Shive S, Tan Y, et al. Community-based colorectal cancer intervention in underserved Korean Americans. *Cancer Epidemiol*. 2009;33(5):381-386.
18. Percac-Lima S, Grant RW, Green AR, et al. A culturally tailored navigator program for colorectal cancer screening in a community health center: a randomized, controlled trial. *J Gen Intern Med*. 2009;24(2):211-217.
19. US Preventive Services Task Force. *Screening for Colorectal Cancer: Recommendation Statement*. Rockville, MD: Agency for Healthcare Research and Quality; 2008.
20. Agency for Healthcare Research and Quality. *2002 AHRQ Practice-Based Research Network (PBRN)*. Rockville, MD: Agency for Healthcare Research and Quality; 2006.
21. Gross CP, McAvay GJ, Krumholz HM, Paltiel AD, Bhasin D, Tinetti ME. The effect of age and chronic illness on life expectancy after a diagnosis of colorectal cancer: implications for screening. *Ann Intern Med*. 2006;145(9):646-653.
22. Harvard Center for Cancer Prevention and the Massachusetts Colorectal Cancer Working Group. *Take Control: Get Tested for Colorectal Cancer*. Boston: Massachusetts Dept of Public Health; 2003.
23. Prochaska JO, DiClemente CC. Toward a comprehensive model of change. In: Miller WR, Healther N, eds. *Treating Addictive Behaviours: Processes of Change*. New York, NY: Plenum Press; 1986:3-27.
24. Dietrich AJ, Tobin JN, Cassells A, et al. Telephone care management to improve cancer screening among low-income women. *Ann Intern Med*. 2006;144(8):563-571.
25. Holden DJ, Jonas DE, Porterfield DS, Reuland D, Harris R. Systematic review: enhancing the use and quality of colorectal cancer screening. *Ann Intern Med*. 2010;152(10):668-676.

26. Jandorf L, Gutierrez Y, Lopez J, Christie J, Itzkowitz SH. Use of a patient navigator to increase colorectal cancer screening in an urban neighborhood health clinic. *J Urban Health*. 2005;82(2):216-224.
27. Neugut AI, Lebwohl B. Screening for colorectal cancer: the glass is half full. *Am J Public Health*. 2009;99(4):592-594.
28. Walsh JME, Salazar R, Nguyen TT, et al. Healthy colon, healthy life: a novel colorectal cancer screening intervention. *Am J Prev Med*. 2010;39(1):1-14.
29. Centers for Disease Control and Prevention. The Community Guide: Task Force Findings: January 1, 1997, to December 14, 2010. The Community Guide Web site. http://www.thecommunityguide.org/about/Conclusion_Report_121410.pdf. Accessed March 22, 2011.
30. Hawley ST, Volk RJ, Krishnamurthy P, Jibaja-Weiss M, Vernon SW, Kneuper S. Preferences for colorectal cancer screening among racially/ethnically diverse primary care patients. *Med Care*. 2008;46(9)(suppl 1):S10-S16.
31. Ling BS, Trauth JM, Fine MJ, et al. Informed decision-making and colorectal cancer screening: is it occurring in primary care? *Med Care*. 2008;46(9)(suppl 1):S23-S29.
32. Hogg WE, Bass M, Calonge N, Crouch H, Satenstein G. Randomized controlled study of customized preventive medicine reminder letters in a community practice. *Can Fam Physician*. 1998;44:81-88.
33. Vinker S, Nakar S, Rosenberg E, Kitai E. The role of family physicians in increasing annual fecal occult blood test screening coverage: a prospective intervention study. *Isr Med Assoc J*. 2002;4(6):424-425.
34. Fang CY, Ma GX, Tan Y, Chi N. A multifaceted intervention to increase cervical cancer screening among underserved Korean women. *Cancer Epidemiol Biomarkers Prev*. 2007;16(6):1298-1302.
35. Agency for Healthcare Research and Quality. Patient Centered Medical Home Resource Center. Agency for Healthcare Research and Quality Web site. http://www.pcmh.ahrq.gov/portal/server.pt/community/pcmn_home/1483/what_is_pcmh_. Accessed April 1, 2011.

INVITED COMMENTARY

Building Blocks of the Patient-Centered Medical/Health Home

The search for the patient-centered medical home resembles the quest for the Holy Grail. Yet no one is sure what this sacred vessel actually looks like. Elements of the patient-centered medical home have been identified by the National Committee on Quality Assurance requirements and meaningful use components from the US federal government (**Table**). Three articles in this issue of the *Archives* provoke thoughts regarding what primary care transformation—the Holy Grail—is all about.

In the study by Lasser et al,[1] patient navigators, who also could be called health care coaches, are community health care workers trained to guide patients through the health care system. Lasser et al found that patients contacted by language-concordant patient navigators are more likely than patients undergoing usual care to be screened for colorectal cancer.

See also pages 897 and 903

Romano and Stafford[2] analyzed 2 ambulatory care databases to examine the association between the use of the electronic health record (EHR)—with or without decision support tools—and clinical quality measures. Decision support tools are reminders to order medications or screening tests recommended by practice guidelines. The study found virtually no difference in 20 quality indicators among visits with no EHR, with an EHR, and with EHR plus decision support.

Commenting on the article by Romano and Stafford, McDonald and Abhyankar [3] suggest that for computerization to improve quality, EHRs need to include clinical decision tools. However, even without decision tools, an EHR combined with a clinical registry can be used to improve clinical care. The EHR-registry component can be used to generate a list of all patients in a medical practice with particular conditions (eg, diabetes mellitus), along with accompanying data (eg, hemoglobin A_{1c} [HbA_{1c}] levels). This information enables practices to identify patients who are overdue for needed interventions (eg, ocular examinations) or who have not achieved consistent disease control (eg, elevated HbA_{1c} levels). Also, registries can generate lists of patients who are overdue for preventive care (eg, mammograms).

How do these 3 articles assist us in understanding the building blocks of transformed primary care, which include the EHR, patient registry, and team-based care? The study by Romano and Stafford and the response by McDonald and Abhyankar demonstrate that the EHR is necessary but not sufficient for improved quality of care. Decision support reminders failed to improve quality of care because physicians must implement the actions that reminders prompt them to perform, for which physicians often do not have adequate time. Treating a typical panel

Table. Elements of the Patient-Centered Medical Home

General Elements
Prompt access to care
Continuity of care
Comprehensive care (care of the whole person and family)
Coordination of care with specialists, hospitals, and other health care services
Controlling the costs of care
Team-based care
Patient-centered care
Additional Elements From the National Committee on Quality Assurance
Registries and panel management
Care management
Self-management support for patients with chronic conditions
Electronic prescribing
Tracking of laboratory and imaging studies and referrals
Reporting on clinical quality measures
Additional Elements From the Meaningful Use Components
Computerization
Up-to-date problem and medication lists
Decision support
Electronic reminder systems

Safe Opioid Prescribing: A Long Way to Go

Jane M. Liebschutz, MD MPH and Daniel P. Alford, MD MPH

Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston Medical Center, Boston, MA, USA.

J Gen Intern Med 26(9):951–2
DOI: 10.1007/s11606-011-1797-3

Chronic pain is one of the most common reasons why patients visit a physician; yet, physicians are poorly trained to manage it. Despite centuries-long therapeutic use of opioids for pain, they fell out of favor for most of the 20th century in the US. However, in the 1980s, medical literature began to support the practice of prescribing opioids for chronic non-cancer pain.[1] Opioid prescribing then escalated over the past two decades,[2] despite lack of strong evidence supporting this practice.[3] Although it started as an antidote to under-treatment of pain, increased opioid prescribing has paralleled increases in opioid misuse including overdoses, diversion and addiction.[4,5]. Over the last decade, the medical and public health communities have begun to address this problem by cautioning that all patients prescribed opioids carry some degree of risk for opioid misuse. Recent clinical guidelines based on expert consensus now endorse universal assessment for opioid misuse risk and monitoring for subsequent potential harm.[6–8] The guidelines suggest that patients at "high risk" for misuse should be identified through individual patient risk factors, such as an existing diagnosis of substance use disorder (SUD), mental health problems, history of legal problems and younger age. Treating chronic pain with opioids should include monitoring for harm including urine drug testing to detect any illicit substance use and whether the prescribed opioid is being taken , frequent prescriber visits, pill counts to evaluate adherence with opioid dosing and to minimize diversion, use of state prescription drug monitoring program data, when available, and addressing aberrant opioid taking behaviors[9,10] such as patients requests for early refills. More intensive monitoring is recommended for those with more risk factors. When necessary, physicians should also refer patients to adjunctive mental health or addiction treatment specialists for co-management, when such services are available.

Few data guide the questions on quantity and frequency of monitoring to identify harm or misuse of opioid analgesics for chronic pain. The guidelines suggest conducting urine drug testing every three to six months for low-risk stable patients, and as often as weekly for high-risk patients.[8] In this journal, Starrels[11] and Morasco[12] use administrative data to examine adherence rates to minimal guideline-based care for patients in two different settings: primary care practices within the University of Pennsylvania Health System and Veterans Affairs Health Centers in the Pacific Northwest. Both authors show that although patients with the identified risk factor of SUD are monitored 2–3 times more closely with urine drug testing than patients without identified SUD, only a minority of these high risk patients are being tested. Morasco and colleagues found that even with the infrequent testing, a substantial portion of patients tested had illicit substances in the urine, including 1 in 20 patients without SUD and 1 in 7 patients with SUD. Furthermore, data from Starrels and colleagues suggest that patients with identified SUD received more frequent early prescriptions on a monthly basis including 1 out of 4 receiving multiple early refills over the duration of time receiving chronic opioids (mean 1.9 years).

Why are monitoring rates in these two systems of care so much lower than that suggested by clinical guidelines? Is it lack of familiarity with recent guidelines, or reluctance to institute time-intensive clinical practices that lack a sufficient evidence base? We believe that several factors may contribute to non-adherence to guidelines. There is a well known lack of formal pain and addiction curricula in medical schools, residencies and in continuing medical education.[13,14] In response to these gaps in medical education, federal agencies have started investing in prescriber education. In 2007 the National Institute on Drug Abuse (NIDA) partnered with eight US medical schools to develop innovative drug abuse and addiction curriculum resources.[15] Recently the FDA met with members of the pharmaceutical industry to outline new requirements for manufacturers of certain opioid analgesics to develop Risk Evaluation and Mitigation Strategies (REMS) that will require physician education.[16] As an additional sign of increased attention, in fiscal year 2010, the National Institutes of Health designated research on prescription drug abuse one of its 220 research topic funding priorities, at which time $36 million was expended in this area.

Even if the rate of monitoring using urine drug testing can be increased, it is not clear that monitoring alone can improve clinical care and thus decrease the rate of opioid misuse.[17] For example, near-universal implementation of a visual analog pain scale in the VA health system nationally was not associated with improvement in clinical care for pain.[18] Clinical care improvement requires education about interpretation of the assessment or monitoring tests, as well as effective communication skills in talking to patients about concerns about opioid misuse. Urine drug testing in particular is quite complex, and requires knowledge of the opioid chemical derivative (e.g. synthetic, semi-synthetic or naturally occurring) and potential metabolites, the duration of detection of the opioid and time of last ingested dose, and the type of assay performed (liquid chromatography vs. immunoassay) and positive cut-offs used by the lab.[19] Any unexpected result requires appropriate inquiry with the patient to help interpret the findings and place them into context. Furthermore, incorrect interpretation of urine drug testing may result in a

Published online July 7, 2011

patient being inappropriately discontinued from useful therapy and may lead to a subsequent rift in trust between patient and physician.

Physicians' attitudes toward opioid prescribing vary widely, ranging from no prescribing at all to very liberal prescribing.[20,21] There are geographic variations in both prescribing and abuse of prescription opioid analgesics.[22,23] Thus, the increasing prevalence of misuse of and addiction to prescription opioid analgesics attributable to physician prescription appears to be the result of a perfect storm: inconsistent and inadequate physician education, lack of sufficient evidence of efficacy and safety of opioid analgesia for chronic pain, and lack of adherence to guideline-based risk assessment and monitoring.

A multifaceted approach to improve opioid prescribing efficacy and safety is urgently needed. Such an approach must start with clinical policies to monitor all patients, while basing the intensity of monitoring on the individual patient's opioid misuse risk. It will need to include systems approaches, such as use of the electronic health record (EHR) to track a patient's adherence to guideline-based treatment plans. Using automated systems, point of care clinical decision support tools and statewide prescription drug monitoring program data may not only lessen the burden on physicians but may provide higher quality care.[22] It is critical that primary care physicians collaborate with colleagues in behavioral health, pharmacy, toxicology and specialty addiction and pain medicine to share knowledge and consultation on prescribing, monitoring and treatment plans if a patient is developing problems (e.g., addiction). This necessary collaboration may be facilitated in the setting of patient-centered medical homes. Additionally, patients need education on the limitations of opioid analgesics, appreciation of the risks associated with opioids (e.g., overdose) and ways to mitigate any risks (e.g. no increases in opioid dose in between visits). Research must be conducted in real world clinical settings where a multifaceted systems approach is examined for effectiveness in improving both individual and population-based outcomes.

Corresponding Author: *Jane M. Liebschutz, MD MPH; Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA (e-mail: jane.liebschutz@bmc.org).*

REFERENCES

1. **Portenoy RK, Foley KM.** Chronic use of opioid analgesics in non-malignant pain: Report of 38 cases. Pain. 1986;25(2):171–86.
2. **Olsen Y, Daumit GL, Ford DE.** Opioid prescriptions by U.S. primary care physicians from 1992 to 2001. J Pain. 2006;7(4):225–35.
3. **Martell BA, O'Connor PG, Kerns RD, et al.** Systematic review: Opioid treatment for chronic back pain: Prevalence, efficacy, and association with addiction. Ann Intern Med. 2007;146(2):116–27.
4. **Paulozzi LJ, Budnitz DS, Xi Y.** Increasing deaths from opioid analgesics in the united states. Pharmacoepidemiol Drug Saf. 2006;15(9):618–27.
5. **Paulozzi LJ, Xi Y.** Recent changes in drug poisoning mortality in the united states by urban-rural status and by drug type. Pharmacoepidemiol Drug Saf. 2008;17(10):997–1005.
6. Model policy for the use of controlled substances for the treatment of pain. Available at: http://www.fsmb.org/pain-model-policy.html. Accessed June 19, 2011.
7. **Gourlay DL, Heit HA, Almahrezi A.** Universal precautions in pain medicine: A rational approach to the treatment of chronic pain. Pain Med. 2005;6(2):107–12.
8. **Chou R, Fanciullo GJ, Fine PG, et al.** Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. J Pain. 2009;10 (2):113–30.
9. **Passik SD, Kirsh KL.** Assessing aberrant drug-taking behaviors in the patient with chronic pain. Current Pain & Headache Reports. 2004;8 (4):289–94.
10. **Passik SD, Messina J, Golsorkhi A, Xie F.** Aberrant drug-related behavior observed during clinical studies involving patients taking chronic opioid therapy for persistent pain and fentanyl buccal tablet for breakthrough pain. J Pain Symptom Manage. 2011;41(1):116–25.
11. **Starrels JL, Becker WC, Weiner MG, Li X, Heo M, Turner BJ.** Low use of opioid risk reduction strategies in primary care even for high risk patients with chronic pain. Journal of General Internal Medicine. 2011: doi:10.1007/s11606-011-1648-2.
12. **Morasco BJ, Duckart JP, Dobschz SK.** Adherence to clinical guidelines for opioid therapy for chronic pain in patients with substance use disorder. Journal of General Internal Medicine. 2011: doi:10.1007/s11606-011-1734-5.
13. **Yanni LM, McKinney-Ketchum JL, Harrington SBH, C, et al.** Preparation, confidence and attitudes about chronic noncancer pain in graduate medical education. J Grad Med Ed. 2010:260–8.
14. **O'Connor PG, Nyquist JG, McLellan AT**. Integrating addiction medicine into graduate medical education in primary care: The time has come. Ann Intern Med. 2011;154(1):56–9.
15. NIDA Centers of Excellence. Available at: http://www.drugabuse.gov/coe/, June 14, 2011.
16. **Okie S.** A flood of opioids, a rising tide of deaths. N Engl J Med. 2010;363 (21):1981–5.
17. **Starrels JL, Becker WC, Alford DP, Kapoor A, Williams AR, Turner BJ.** Systematic review: Treatment agreements and urine drug testing to reduce opioid misuse in patients with chronic pain. Ann Intern Med. 2010;152(11):712–20.
18. **Mularski RA, White-Chu F, Overbay D, Miller L, Asch SM, Ganzini L.** Measuring pain as the 5th vital sign does not improve quality of pain management. J Gen Intern Med. 2006;21(6):607–12.
19. **Heit HA, Gourlay DL.** Urine drug testing in pain medicine. J Pain Symptom Manage. 2004;27(3):260–7.
20. **Bhamb B, Brown D, Hariharan J, Anderson J, Balousek S, Fleming MF.** Survey of select practice behaviors by primary care physicians on the use of opioids for chronic pain. Curr Med Res Opin. 2006;22 (9):1859–65.
21. **Scanlon MN, Chugh U.** Exploring physicians' comfort level with opioids for chronic noncancer pain. Pain Res Manag. 2004;9(4):195–201.
22. **Curtis LH, Stoddard J, Radeva JI, et al.** Geographic variation in the prescription of schedule II opioid analgesics among outpatients in the united states. Health Serv Res. 2006;41(3 Pt 1):837–55.
23. **Webster BS, Cifuentes M, Verma S, Pransky G.** Geographic variation in opioid prescribing for acute, work-related, low back pain and associated factors: A multilevel analysis. Am J Ind Med. 2009;52(2):162–71.

Proton Pump Inhibitor Discontinuation in Long-Term Care

Amy Linsky, MD, MSc,[*†] *John A. Hermos, MD,*[*‡] *Elizabeth V. Lawler, DSc,*[‡§] *and James L. Rudolph, MD, SM*[‖§]

OBJECTIVES: To determine factors associated with proton pump inhibitor (PPI) discontinuation in long-term care.

DESIGN: Retrospective cohort analysis.

SETTING: Veterans Affairs (VA) long-term care facilities.

PARTICIPANTS: Veterans admitted for nonhospice care in 2005 with a length of stay of 7 days or more who were prescribed a PPI within 7 days of admission (N = 10,371).

MEASUREMENTS: Prescribed medications and comorbidities were determined from VA pharmacy and administrative databases and functional status from Minimum Data Set records. Associations between participant characteristics and PPI discontinuation were determined using Cox proportional hazard ratios (HRs), censoring at death, discharge, or 180 days after admission.

RESULTS: Participants were predominantly male (97%) and had a median age of 73 (interquartile range 60–81). There were 2,749 (27%) PPI discontinuations; 43% of these occurred within 28 days of admission. Hospitalizations (HR = 1.22, 95% confidence interval (CI) = 1.01–1.46), preadmission PPI use (HR = 1.35, 95% CI = 1.16–1.56), and lowest functional status (HR = 1.22, 95% CI = 1.03–1.45) were associated with early PPI discontinuation in adjusted models. Participants with gastric acid–related disease (HR = 0.53, 95% CI 0.46–0.61), diabetes mellitus (HR = 0.82, 95% CI 0.72–0.94), and those who were prescribed six or more medications (6–7 medications, HR = 0.78, 95% CI = 0.66–0.92; 8–10 medications, HR = 0.64, 95% CI = 0.54–0.76; ≥11 medications 0.51, 95% CI = 0.42–0.62) were less likely to have early discontinuation. No PPI discontinuer had PPIs resumed during the study, and few (9%) had histamine-2 receptor antagonist substitutions.

CONCLUSION: Although there may be clinical uncertainty regarding PPI discontinuation, more than one-quarter of participants prescribed a PPI upon admission to long-term care had it discontinued within 180 days. Targeting individuals prescribed PPIs for medication appropriateness review may reduce prescribing of potentially nonindicated medications. **J Am Geriatr Soc 59:1658–1664, 2011.**

Key words: long-term care; polypharmacy; proton pump inhibitors; prescriptions

Proton pump inhibitors (PPIs), a class of medications frequently prescribed to residents of long-term care facilities, are used to suppress production of gastric acid. In the short term, PPIs are highly effective for treating gastric acid–related diseases such as gastroesophageal reflux, esophagitis, and gastric and duodenal ulcers. Although there are critical indications for long-term use (e.g., Barrett's esophagus), chronic use is often not indicated.[1] PPI use without clear indication is common in people in hospitals,[2,3] and regardless of appropriateness of inpatient use, these individuals are then frequently discharged with a PPI prescription.[4] Although generally considered safe, recent observational studies have associated PPIs with risks of community-acquired and hospital-acquired pneumonia, *Clostridium difficile* infection, and osteoporotic fractures.[5–9] The associated morbidity of PPIs combined with the fragility of the long-term care population suggests that overuse of this medication class may disproportionately negatively affect these individuals.

Discontinuation of unnecessary or ineffective medications is challenging in long-term care settings, where residents take an average of seven to eight medications daily[10] and are at higher risk than older community-dwelling adults for adverse drug events and other medication-related complications.[11,12] However, the transition to a long-term care facility provides an opportunity to review and change prescribed medications under supervised conditions.[13] Considering the importance of these transitional periods and the fact that residents in long-term care are at high risk

From the [*]Section of General Internal Medicine, Department of Medicine, School of Medicine, Boston University, Boston, Massachusetts; [†]Center for Organization, Leadership and Management Research, [‡]Massachusetts Veterans Epidemiology Research and Information Center, Veterans Affairs Cooperative Studies Program, and [‖]Geriatric Research, Education and Clinical Center, Veterans Affairs Boston Healthcare System, Boston, Massachusetts; and [§]Division of Aging, Department of Medicine, Harvard Medical School, Boston, Massachusetts.

Address correspondence to Amy Linsky, 801 Massachusetts Avenue, 2nd Floor, GIM, Boston, MA 02118. E-mail: Amy.Linsky@bmc.org

DOI: 10.1111/j.1532-5415.2011.03545.x

0002-8614/11/$15.00

for adverse drug events,[14] the objectives of this study were to characterize the discontinuation of PPIs for residents newly admitted to Veterans Affairs (VA) long-term care facilities, identify factors independently associated with PPI discontinuation, and determine whether histamine$_2$ receptor antagonist (H_2RA) initiation for continued gastric acid suppression accompanied PPI discontinuation. It was hypothesized that discontinuation would occur early in the admission, that a greater number of total medications would be associated with PPI discontinuation, and that few residents would have a substitution of medication class from PPIs to H_2RAs. To accomplish these objectives, a retrospective epidemiological study of residents admitted to VA long-term care facilities was conducted using administrative, pharmacy, and Minimum Data Set (MDS) data.

METHODS

Study Setting and Data Sources

National VA administrative databases linked with MDS records were used to evaluate PPI prescribing in VA long-term care facilities. The MDS is a standardized assessment of nursing home residents that documents demographic information, functional status in terms of activities of daily living, and healthcare system use.[15] The Veterans Health Administration mandates completion upon admission and discharge and approximately every 3 months during any long-term care admission for which federal money is received. The VA Decision Support System (DSS) and National Patient Care Database (NPCD) files contain information on admission and discharge dates, age, sex, medical comorbidities, and pharmacy dispensing records.

Analyses were conducted at the Massachusetts Veterans Epidemiology Research and Information Center, VA Cooperative Studies Coordinating Center, VA Boston Healthcare System. The institutional review board of the VA Boston Healthcare System approved the study.

Study Population

Ten thousand eight hundred thirty-eight veterans newly admitted to a VA long-term care facility in 2005 and meeting the following criteria were identified: a completed admission MDS assessment, an admission for neither hospice nor respite care, and a prescription for a PPI within 7 days of admission. Only the first admission in 2005 per resident was selected, and 1 calendar year was chosen to minimize potential chronological trends in PPI prescribing. The study population was restricted to 10,624 individuals with a matching admission date in the VA Extended Care files of the NPCD, and then 234 individuals with a length of stay less than 7 days were excluded to allow for adequate opportunity to discontinue medications; 19 individuals without a complete functional assessment were further excluded, creating a cohort of 10,371 residents in 129 long-term care facilities analyzed in this study. Length of stay was determined according to discharge date, with a maximum follow-up of 180 days. Discharge status included discharge home, transfer to other facilities, and death. These facilities were renamed Community Living Centers in 2008 but will be referred to herein by the broader descriptive term long-term care facilities.

PPI Exposure

In the VA, long-term care residents have medications dispensed using procedures similar to those used for patients who are hospitalized. Residents are not permitted to take their own outpatient or over-the-counter medications, medications are individually bar-coded for tracking, and nurses dispense all medications. From DSS data, all days during which PPIs were dispensed to patients were determined, and the last day of PPI use was calculated by adding the days supplied to the final date dispensed.

Baseline PPI Exposure

Residents were designated as baseline users if there was any dispensing of oral PPIs in the first 7 days of the long-term care admission. This window allowed providers time for patient evaluation, medication reconciliation, and order entry. Day 7 of admission, the end of the defined enrollment period, was then set as the first day of use for the observation period.

PPI Duration

Duration of use was calculated as the number of days from Day 7 until the date of last PPI use because of discontinuation or censoring. For residents whose PPI was discontinued before Day 7, their duration was calculated by counting the days between the first date of PPI dispensing in the long-term care facility and the last day of PPI use. All measured durations of PPI use were continuous because no participant had intervening periods without PPI prescriptions.

PPI Discontinuation

If the calculated last day of PPI use was before discharge from the long-term care facility, then the participant was designated as a discontinuer. If the last day of use was at or extended beyond discharge or 180 days, the participant was designated as a continuer. No discontinuer had a subsequent reinitiation of PPI during the study admission.

Prior PPI Exposure

PPI use in the 30 days before long-term care admission was determined from VA DSS inpatient and outpatient pharmacy records. If participants had a PPI prescription with a calculated end date at or within 30 days before long-term care admission, or if they had acute care inpatient dispensing during that time, they were considered to be preadmission PPI users. Those without such documented PPI dispensing were classified as nonusers during this preadmission period. It was determined a priori that PPI exposure that ceased more than 30 days before long-term care admission probably had no remaining physiological effect.[16]

Other Medication Exposure

Histamine-2 Receptor Antagonist Use

Histamine-2 receptor antagonists (H_2RAs) have indications similar to those of PPIs, but they are neither as potent nor typically as long acting in their suppressive effect on gastric acid secretion. Because H_2RAs may be used alternatively to, or less commonly concurrently with, PPIs for gastric acid suppression, VA pharmacy records were queried for H_2RA prescriptions for the 10,371 study participants to determine whether they had a substitution in class of acid-suppressing

medication as opposed to a simple discontinuation. If participants had a H_2RA prescription within the first 7 days of long-term care admission, they were classified as baseline H_2RA users (n = 533). For those not prescribed baseline H_2RAs (n = 9,838), whether and when it was initiated during their admission was determined. In this way, six subgroups were created based on combinations of PPI use (discontinuation vs continuation) and H_2RA use (baseline vs initiation vs no use). For the subgroup comprising participants with PPI discontinuation and H_2RA initiation (a substitution of gastric acid–suppressing therapy), it was determined when, relative to PPI discontinuation, the H_2RA initiation occurred (>1 week before, ≤1 week before, ≤1 week after, >1 week after).

Medication Use Other Than Gastric Acid Suppressants

The number and type of medications that individuals take typically reflect medical and psychiatric comorbidities.[17] The number of distinct oral, intravenous, and topical medications, excluding PPIs and H_2RAs, was counted. The count could include medications prescribed on an as-needed basis if they were dispensed on the date of interest because the count reflects pharmacy dispensing and not computer order entry. Because Day 7 of admission was the first day of the analysis period, the number of medications on that day was categorized into quartiles for use in regression analyses.

Because the concurrent use of specific medications may serve as an indication for continuation of PPIs, the prevalence of use on Day 7 of aspirin, antiplatelet agents other than aspirin, nonsteroidal anti-inflammatory drugs (NSAIDs), and systemic glucocorticoids was determined. Aspirin was further separated into low dose (81 mg) and other dosing for presumed differences in potential gastric irritation.

Other Variables

Participant Demographics and Medical Conditions

Age at admission and sex were obtained from administrative databases. Medical comorbidities diagnosed in the year before admission or during the study window were also recorded. Specific comorbidities and their *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) codes are as follows: gastric acid–related disorders (including diseases of the esophagus (530), gastric ulcer (531), duodenal ulcer (532), and gastritis and duodenitis (535)), dementia (290), diabetes mellitus (250), essential hypertension (401), chronic obstructive pulmonary disease (490–496), arthritis (711–715), chronic liver disease and cirrhosis (571), and chronic kidney disease (585–586).

Functional Status

Functional status indicators have been shown to predict mortality and have been associated with the presence of unnecessary medications.[18,19] From admission MDS records, individual performances on seven distinct activities of daily living (ADLs; dressing, personal hygiene, toilet use, locomotion on unit, transfer, bed mobility, and eating) were summed to create a composite score, the ADL Long Form.[20]

Performance on each ADL ranged from 0 (independent) to 4 (total dependence). If the activity had not occurred in the prior 7 days, the participant was considered to have total dependence for that activity and scored accordingly. Composite scores range from 0 to 28 and were categorized into quartiles for use in further analyses. There are no consensus recommendations for the analytic use of the ADL Long Form.

Hospitalizations

MDS admission assessments document the number of hospitalizations to VA Medical Centers or community facilities that occurred in the 90 days before long-term care admission. Hospitalizations were categorized as none, one, or two or more.

Analyses

Participant Characteristics

Participant age was quantified using median and interquartile range (IQR) and participant sex, medical comorbidities, preadmission PPI use, and specific medication exposures using prevalence. The medication count on admission Day 7 and the composite ADL score were each categorized into quartiles.

PPI Discontinuation and Factors Associated with PPI Discontinuation

The unadjusted prevalence of PPI discontinuation during the study was calculated. Actuarial methods were then used to derive unadjusted survival curves demonstrating the time to discontinuation to account for the discrete opportunities for discontinuation. This analysis also yielded the hazards of discontinuation throughout the study.

Characteristics which were associated with time to discontinuation were determined using Cox proportional hazard regressions. Separate models were run for individual comorbidities, medication count quartiles, specific medication exposures, ADL quartiles, hospitalizations, preadmission PPI use, baseline H_2RA use, age, and sex. Variables with moderate significance ($P<.10$) and variables with clinical relevance (age, sex, dementia) were entered into a multivariable model, using backward selection based on statistical significance to reach a final model. Sensitivity analyses compared the effect of modeling each of the four gastric acid–related disease categories separately with the effect of modeling them as a combined variable.

The proportional hazards assumption was tested by including interaction terms between each variable and the natural log of time. Because the assumption was violated for several variables, and given the nature of the hazard plot, piecewise analyses identified factors associated with discontinuation within the first 28 days of admission (21 days of PPI duration) and those associated with discontinuation after Day 28.

Patterns of PPI and H_2RA Prescribing

The prevalence of baseline H_2RA use, H_2RA initiation, and no H_2RA exposure were each determined for PPI discontinuers and continuers. For participants with PPI discontinuation and H_2RA initiation (a substitution), the proportion of participants who had H_2RAs started at various times before and after PPI discontinuation was calculated.

All analyses were conducted using SAS version 9.1 (SAS Institute, Inc., Cary, NC). Unless otherwise specified, alpha = 0.05.

RESULTS

The study population of 10,371 VA long-term care residents with an admission in 2005 and PPI prescription dispensed within the first 7 days of admission was predominantly male (97%) and had a median age of 73 (IQR 60–81) (Table 1). In this group of baseline PPI users, there was a high prevalence of comorbid conditions, including 46% with one or more ICD-9-CM–documented gastric acid–related diseases, most of which were gastroesophageal reflux or reflux esophagitis (n = 4,027, 39%). The majority (67%) had a record of a PPI prescription in the 30 days before admission. Nearly two in three participants had at least one hospitalization in the 90 days before admission. The median number of medications on admission Day 7, not including PPIs and H_2RAs, was 7 (IQR 5–10). Nearly one in three residents had exposure to low-dose (n = 1,989, 19%) or other dose (n = 1,185, 11%) aspirin, and an additional 902 residents (9%) had exposure to other antiplatelet agents. The median score for the summative seven-item ADL Long Form was 8 (IQR 2–15), suggesting moderate to severe functional impairment.

Table 1. Population Baseline Characteristics of Veterans Affairs Long-Term Care Patients Receiving a Proton Pump Inhibitor (PPI) within 7 Days of Admission (N = 10,371)

Characteristic	Value
Age, median (IQR)	73 (60–81)
Male, n (%)	10,043 (96.8)
Comorbidities (365 days before admission), n (%)	
Dementia	929 (9.0)
Diabetes mellitus	4,438 (42.8)
Gastric acid–related diseases*	4,815 (46.4)
Diseases of the esophagus	4,353 (42.0)
Gastric ulcer	187 (1.8)
Duodenal ulcer	177 (1.7)
Gastritis and duodenitis	689 (6.6)
Chronic obstructive pulmonary disease	4,624 (44.6)
Chronic kidney disease	1,623 (15.7)
Chronic liver disease	676 (6.5)
Hypertension	8,170 (78.8)
Arthritis	3,669 (35.4)
Physical function quartile (activity of daily living long form score), n (%)†	
1 (0–2) highest function	3,038 (29.3)
2 (3–8)	2,537 (24.5)
3 (9–15)	2,369 (22.8)
4 (16–28) lowest function	2,427 (23.4)
Hospitalizations in 90 days before admission, n (%)†	
0	3,608 (34.8)
1	4,924 (47.5)
≥2	1,839 (17.7)
PPI dispensed in 30 days before admission, n (%)	6,913 (66.7)
Number of medications on admission Day 7, median (IQR) (n = 9,806)‡	7 (5–10)
Specific medication exposure on admission Day 7, n (%)§	
Aspirin, 81 mg	1,989 (19.2)
Aspirin, other than 81 mg	1,185 (11.4)
Antiplatelet agents other than aspirin	902 (8.7)
Nonsteroidal anti-inflammatory drug	467 (4.5)
Systemic glucocorticoid	1,182 (11.4)

* Gastric acid–related diseases are not mutually exclusive.

† Data obtained from admission assessment of the Minimum Data Set.

‡ Number of medications is exclusive of PPIs and histamine = 2 receptor antagonists.

§ Aspirin 81 mg, aspirin other dose, and antiplatelet agents are mutually exclusive.

IQR = interquartile range.

PPI Discontinuation and Factors Associated with PPI Discontinuation

Slightly more than one in four participants had their PPI prescription discontinued before the end of the study period (n = 2,749, 27%). PPI discontinuation was more likely early in the observation period, with many of the discontinuations occurring in the first 28 days (n = 1,173, 43% of discontinuations). This is demonstrated in the survival curve and associated hazard rate curve (Figure 1A and B).

In unadjusted Cox proportional hazard models, factors found to be statistically significant predictors of PPI discontinuation were two or more hospitalizations in the 90 days before admission to the long-term care facility, preadmission PPI use, and being in the lowest quartile of functional status (Table 2). Factors associated with lower hazard of PPI discontinuation were six or more medications on admission Day 7; exposure to low-dose aspirin or antiplatelet agents on admission Day 7; and diagnoses of gastric acid-related disease, diabetes mellitus, chronic obstructive pulmonary disease, and hypertension.

In piecewise, adjusted analyses stratified according to time, factors associated with PPI discontinuation before Day 28 of admission (early) were hospitalization, preadmission PPI use, and worse physical function (Table 2). Gastric acid–related disease, diabetes mellitus, and six or more medications were associated with a lower hazard of PPI discontinuation before Day 28 of admission. At or after 28 days (late), only two or more hospitalizations before admission was associated with greater hazard of PPI discontinuation, whereas gastric acid–related disease remained associated with lower hazard.

Sensitivity analyses including discrete gastric acid diagnoses found estimates of association with PPI discontinuation similar to analyses with a combined variable. The hazard ratios for early discontinuation ranged from 0.45 to 0.77, compared with 0.53 for the combined variable, whereas those for late discontinuation ranged from 0.75 to 0.86, compared with 0.80 for the combined variable.

Patterns of PPI and H_2RA Prescribing

In participants with eventual PPI discontinuation (n = 2,749), 2,328 (85%) had no H_2RA exposure during the study period, 179 (6%) received an H_2RA at baseline, and 242 (9%) had an H_2RA initiated during the study window. Of these 242 H_2RA initiators, most had this medication first prescribed within 7 days before PPI

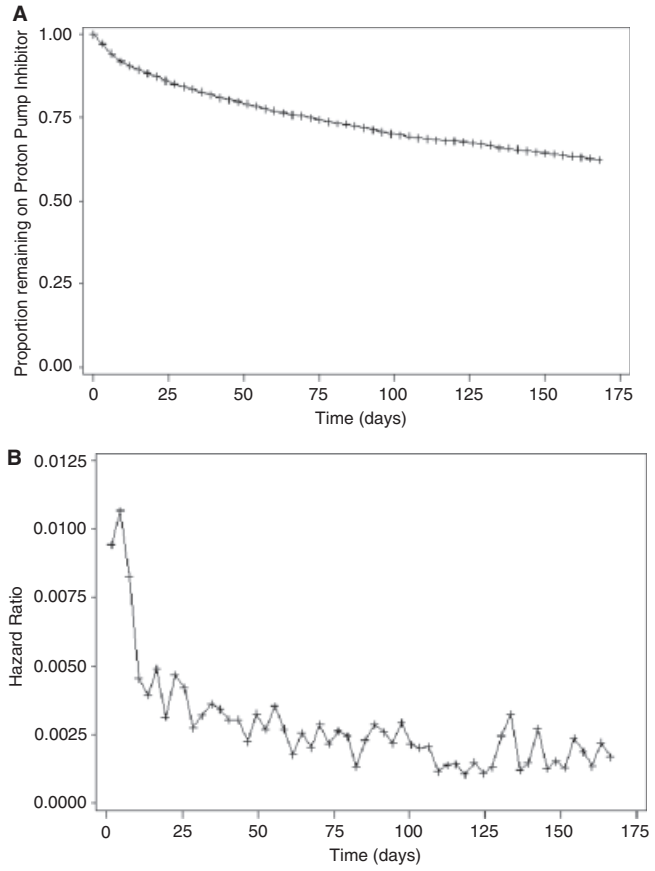


Figure 1. Time to proton pump inhibitor (PPI) discontinuation in individuals with baseline prescriptions at admission to long-term care. Survival and hazard curves for time to PPI discontinuation. (A) Time to PPI discontinuation; note the steep discontinuation in the first 21 days (28 days of admission). (B). Associated hazard rate for PPI discontinuation. The start of analysis is Day 7 of admission to long-term care.

discontinuation (n = 104, 43%) or more than 1 week before discontinuation (n = 71, 29%) (Figure 2).

DISCUSSION

This study describes PPI discontinuation and factors associated with discontinuation. Of residents admitted to VA long-term care facilities with PPI prescriptions, 26.5% had that medication discontinued, more than 40% of which occurred within the first 4 weeks of admission. Factors predictive of early discontinuation included a hospitalization in the 90 days before admission, preadmission PPI use, and greater functional impairment. Consistent with indication-based prescribing, participants with gastric acid–related disease were less likely to have PPIs discontinued during the observation window, as were those with diabetes mellitus. PPI discontinuation did not result in subsequent resumption and infrequently led to H_2RA initiation, suggesting that these individuals were able to be managed without significant clinical consequences. Given the frequency with which people are prescribed PPIs without documented indication, these data suggest that many may be able to have this potent gastric acid suppressant safely withdrawn from their medication regimens.

These rates of PPI use and discontinuation are consistent with those from other studies that illustrate overuse of PPIs[21,22] but add to current knowledge by identifying timing of and factors associated with PPI discontinuation. Although there are no benchmark standards for PPI discontinuation, one study found that 27% of community-dwelling individuals receiving long-term PPI therapy were able to withdraw without resuming use for at least 1 year.[23] Long-term care residents may have different requirements for PPI use than acute care inpatients or community-dwelling older adults, although one study found that 61% of participants were prescribed a PPI on transfer to a skilled nursing facility and that only half of these had a diagnosis indicating appropriate use.[24] This is consistent with the finding that only 46% of the study population had documented gastric acid–related disease. Finally, by using a large national database and incorporating time to discontinuation, it was possible to demonstrate that discontinuation more often occurred early in the admission and may reflect active attention to treatment plans during transitions in care.

Because one concern related to medication appropriateness is a balance between time to benefit and expected lifespan, many individuals near the end of life are on medications from which they are unlikely to derive benefit. This study found that participants with worse functional status were more likely to have their PPI discontinued, which may indicate that these frailer, sicker individuals were seen as appropriate candidates for medication review and discontinuation of nonessential prescriptions. Such individuals in this study may also be more likely to have unnecessary medications as part of their regimen and thus have more opportunities for discontinuation.[19] Contrary to the hypothesis, individuals with six or more total medications were less likely to have a PPI discontinued. Because such individuals are at higher risk for adverse drug events, they are a prime population to target for better medication prescribing. Careful review of all medications and elimination of those without current clinical indication could lead to fewer unintended consequences and lower associated healthcare costs.

This study benefited from a large, national sample encompassing extended care admissions from all VA long-term care facilities and the integrated medical record of the VA healthcare system, which enables a cohesive view of a resident's comorbidities, healthcare usage, and medication regimen before and during admission. The demographics of the VA population (largely older men) may limit generalizability to other populations. The standardized medication administration protocols with barcode tracking captures accurate long-term care medication use. Although it was possible to capture dispensing accurately, the lack of data on preadmission use of non-VA prescribed formulations and no specific measure of indication for PPI initiation, continuation, or discontinuation limited this study. Although this is important to determine the appropriateness of PPI use, it was possible to include documented comorbidities and exposure to specific medications of interest, which may serve as a proxy for indication. Potential misclassification in administrative databases limits this.

Table 2. Cox Proportional Hazard Models Predicting Proton Pump Inhibitor (PPI) Discontinuation

	Hazard Ratio (95% Confidence Interval)		
	Unadjusted	Adjusted*	
Variable		Admit Days 0–27	Admit Days 28–180
Age	1.00 (1.00–1.00)		
Male	0.96 (0.78–1.19)		
Comorbidities			
Dementia	0.96 (0.85–1.08)	1.00 (0.80–1.24)	0.92 (0.78–1.08)
Diabetes mellitus	0.86 (0.80–0.93)	0.82 (0.72–0.94)	0.90 (0.81–1.01)
Gastric acid–related disease	0.70 (0.64–0.75)	0.53 (0.46–0.61)	0.80 (0.72–0.89)
Chronic obstructive pulmonary disease	0.89 (0.82–0.96)	0.93 (0.82–1.05)	0.91 (0.82–1.01)
Chronic kidney disease	0.91 (0.82–1.01)		
Chronic liver disease	0.91 (0.77–1.07)		
Hypertension	0.90 (0.83–0.98)	1.07 (0.92–1.25)	0.90 (0.80–1.02)
Arthritis	0.99 (0.91–1.07)		
Physical function quartile (activity of daily living long form score) (reference 1 (0–2))			
2 (3–8)	1.08 (0.97–1.21)	1.07 (0.89–1.28)	1.10 (0.94–1.28)
3 (9–15)	1.05 (0.94–1.17)	1.06 (0.88–1.27)	1.01 (0.87–1.17)
4 (16–28) lowest function	1.21 (1.09–1.34)	1.22 (1.03–1.45)	1.09 (0.94–1.25)
Hospitalizations in 90 days before admission (reference 0)			
1	1.06 (0.98–1.15)	1.17 (1.01–1.35)	1.00 (0.89–1.13)
≥2	1.13 (1.01–1.26)	1.22 (1.01–1.46)	1.18 (1.02–1.38)
PPI use in 30 days pre-admission	1.10 (1.02–1.19)	1.35 (1.16–1.56)	1.01 (0.91–1.13)
Number of medications on admission Day 7 (number of medications)† (n = 9,806) (reference 1 (0–5))			
2 (6–7)	0.88 (0.79–0.99)	0.78 (0.66–0.92)	0.99 (0.86–1.15)
3 (8–10)	0.78 (0.70–0.86)	0.64 (0.54–0.76)	0.94 (0.82–1.07)
4 (≥11)	0.70 (0.63–0.78)	0.52 (0.42–0.63)	0.90 (0.77–1.05)
Specific medications‡			
Aspirin, 81 mg	0.88 (0.80–0.97)	1.01 (0.85–1.21)	1.06 (0.93–1.21)
Aspirin, other than 81 mg	0.95 (0.84–1.07)		
Antiplatelet agents	0.84 (0.73–0.97)	0.95 (0.74–1.22)	1.01 (0.84–1.21)
Nonsteroidal anti-inflammatory drug	0.84 (0.69–1.03)		
Systemic glucocorticoid	0.92 (0.81–1.04)		

* Model adjusted for all variables that were significant in unadjusted analyses.
† Number of medications is exclusive of PPIs and histamine = 2 receptor antagonists.
‡ Aspirin 81 mg, aspirin other dose, and antiplatelet agents are mutually exclusive.

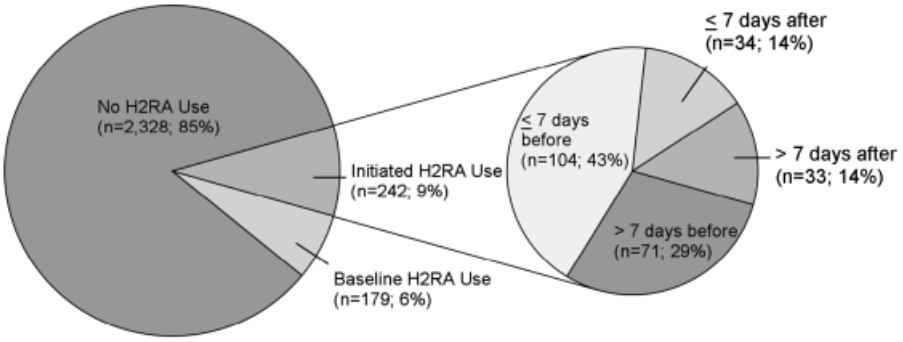


Figure 2. Patterns of histamine-2 receptor antagonist (H_2RA) use in individuals with proton pump inhibitor (PPI) discontinuation. The pie chart (left) shows the proportion of individuals with PPI discontinuation who had no H_2RA prescriptions, had baseline prescriptions, or initiated prescriptions. Few individuals with PPI prescriptions had H_2RA prescriptions during the study. For the subgroup with H_2RA initiation, the pie chart (right) illustrates the timing of the H_2RA initiation referent to the day of PPI discontinuation.

By incorporating data from the MDS, it was possible to assess resident functional status, an important predictor of morbidity and healthcare utilization.[15] Finally, this study included participant characteristics that may be associated with prescribing patterns but was unable to account for potential variation in pharmacist involvement or prescriber characteristics.

PPIs have been demonstrated to be effective for several conditions but are frequently initiated and perpetuated in chronically ill individuals without clear indication. The presence of multiple chronic conditions can lead to development of complicated medication regimens, which in turn may reduce of the ability of providers to isolate prescriptions that are no longer necessary. Therefore, clinical focus on functionally impaired individuals prescribed many medications would be appropriate for targeted medication review, including evaluation of dose, duration, and indication. Given the potential for associated negative consequences, individuals not meeting prescribing indications could reasonably have a trial of PPI withdrawal, despite apparent tolerance of the medication without overt adverse drug events. Individuals transitioning between healthcare settings provide an important opportunity to review medications and discontinue those without indication. Future work will better define which individuals will most benefit from these interventions.

In summary, although there is no current consensus or guideline regarding PPI discontinuation, it was possible to show that more than one-quarter of individuals prescribed a PPI upon admission to long-term care had it discontinued within 180 days. These findings suggest that opportunities exist in long-term facilities to evaluate and reduce prescribing of medications that may be noncritical or inappropriate. Once barriers to medication discontinuation, and to what extent they can be attributed to individual, provider, or system factors, are recognized, interventions can be implemented to improve appropriate prescribing to older adults, with the ultimate goal of reducing adverse health outcomes and associated healthcare costs due to overmedication.

ACKNOWLEDGMENTS

We especially thank Galina Sokolovskaya, MS, for her programming assistance.

Conflict of Interest: The editor in chief has reviewed the conflict of interest checklist provided by the authors and has determined that the authors have no financial or any other kind of personal conflicts with this paper.

Dr. Rudolph is funded by a VA Rehabilitation Career Development Award. This study was supported using the resources and personnel of the Massachusetts Veterans Epidemiology Research and Information Center as part of the VA Cooperative Studies Programs and the facilities and resources of the VA Boston Healthcare System.

Author Contributions: Study concept and design: Linsky, Hermos, Lawler, Rudolph. Acquisition of data: Linsky, Lawler. Analysis and interpretation of data: Linsky, Hermos, Lawler, Rudolph. Drafting the manuscript: Linsky. Critical revision of the manuscript: Linsky, Hermos, Lawler, Rudolph. Obtained funding: Linsky, Hermos, Lawler, Rudolph.

Sponsor's Role: The sponsors had no role in the design or conduct of the study.

Presented in abstract form at the National Meeting of the Society of General Internal Medicine, May 29, 2010, Minneapolis, Minnesota, and the International Conference on Pharmacoepidemiology and Therapeutic Risk Management, August 21, 2010, Brighton, England.

REFERENCES

1. Choudhry MN, Soran H, Ziglam HM. Overuse and inappropriate prescribing of proton pump inhibitors in patients with *Clostridium difficile*-associated disease. QJM 2008;101:445–448.
2. ASHP Therapeutic Guidelines on Stress Ulcer Prophylaxis. ASHP Commission on Therapeutics and approved by the ASHP Board of Directors on November 14, 1998. Am J Health Syst Pharm 1999;56:347–379.
3. Grube RR, May DB. Stress ulcer prophylaxis in hospitalized patients not in intensive care units. Am J Health Syst Pharm 2007;64:1396–1400.
4. Zink DA, Pohlman M, Barnes M et al. Long-term use of acid suppression started inappropriately during hospitalization. Aliment Pharmacol Ther 2005;21:1203–1209.
5. Dial S, Delaney JA, Barkun AN et al. Use of gastric acid-suppressive agents and the risk of community-acquired *Clostridium difficile*-associated disease. JAMA 2005;294:2989–2995.
6. Laheij RJ, Sturkenboom MC, Hassing RJ et al. Risk of community-acquired pneumonia and use of gastric acid-suppressive drugs. JAMA 2004;292: 1955–1960.
7. Yang YX, Lewis JD, Epstein S et al. Long-term proton pump inhibitor therapy and risk of hip fracture. JAMA 2006;296:2947–2953.
8. Linsky A, Gupta K, Lawler EV et al. Proton pump inhibitors and risk for recurrent *Clostridium difficile* infection. Arch Intern Med 2010;170:772–778.
9. Yang YX, Metz DC. Safety of proton pump inhibitor exposure. Gastroenterology 2010;139:1115–1127.
10. Doshi JA, Shaffer T, Briesacher BA. National estimates of medication use in nursing homes: Findings from the 1997 Medicare Current Beneficiary Survey and the 1996 Medical Expenditure Survey. J Am Geriatr Soc 2005;53: 438–443.
11. Sloane PD, Zimmerman S, Brown LC et al. Inappropriate medication prescribing in residential care/assisted living facilities. J Am Geriatr Soc 2002;50:1001–1011.
12. Nebeker JR, Barach P, Samore MH. Clarifying adverse drug events: A clinician's guide to terminology, documentation, and reporting. Ann Intern Med 2004;140:795–801.
13. Coleman EA. Falling through the cracks: Challenges and opportunities for improving transitional care for persons with continuous complex care needs. J Am Geriatr Soc 2003;51:549–555.
14. Field TS, Gurwitz JH, Avorn J et al. Risk factors for adverse drug events among nursing home residents. Arch Intern Med 2001;161:1629–1634.
15. Mor V. A comprehensive clinical assessment tool to inform policy and practice: Applications of the Minimum Data Set. Med Care 2004;42:11150–11159.
16. Sachs G, Shin JM, Howden CW. Review article: The clinical pharmacology of proton pump inhibitors. Aliment Pharmacol Ther 2006;23(Suppl 2):2–8.
17. Schneeweiss S, Seeger JD, Maclure M et al. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. Am J Epidemiol 2001;154:854–864.
18. Keeler E, Guralnik JM, Tian H et al. The impact of functional status on life expectancy in older persons. J Gerontol A Biol Sci Med Sci 2010;65A: 727–733.
19. Hajjar ER, Hanlon JT, Sloane RJ et al. Unnecessary drug use in frail older people at hospital discharge. J Am Geriatr Soc 2005;53:1518–1523.
20. Morris JN, Fries BE, Morris SA. Scaling ADLs within the MDS. J Gerontol A Biol Sci Med Sci 1999;54A:M546–M553.
21. Ryder SD, O'Reilly S, Miller RJ et al. Long term acid suppressing treatment in general practice. BMJ 1994;308:827–830.
22. Jacobson BC, Ferris TG, Shea TL et al. Who is using chronic acid suppression therapy and why? Am J Gastroenterol 2003;98:51–58.
23. Bjornsson E, Abrahamsson H, Simren M et al. Discontinuation of proton pump inhibitors in patients on long-term therapy: A double-blind, placebo-controlled trial. Aliment Pharmacol Ther 2006;24:945–954.
24. Glew CM, Rentler RJ. Use of proton pump inhibitors and other acid suppressive medications in newly admitted nursing facility patients. J Am Med Dir Assoc 2007;8:607–609.

J Cancer Surviv (2011) 5:27–34
DOI 10.1007/s11764-010-0162-6

Lifestyle behaviors in Massachusetts adult cancer survivors

Amy Linsky · Joshua Nyambose · Tracy A. Battaglia

Received: 30 September 2010 / Accepted: 23 November 2010 / Published online: 5 December 2010

Abstract
Introduction Adoption of healthy lifestyles in cancer survivors has potential to reduce subsequent adverse health. We sought to determine the prevalence of tobacco use, alcohol use, and physical inactivity among cancer survivors overall and site-specific survivors.
Methods We performed a cross-sectional analysis of the Massachusetts Behavioral Risk Factor Surveillance System, 2006–2008, and identified 1,670 survivors and 18,197 controls. Specific cancer sites included prostate, colorectal, female breast, and gynecologic (cervical, ovarian, uterine). Covariates included age, gender, race/ethnicity, education, income, marital status, health insurance, and physical and mental health. Gender stratified logistic regression models associated survivorship with each health behavior.
Results 4.9% of men and 7.7% of women reported a cancer history. In adjusted regression models, male survivors were similar to gender matched controls, while female survivors had comparable tobacco and alcohol use but had more physical inactivity than controls (OR 1.5; 95% CI, 1.2–1.8). By site, breast cancer survivors were more likely to be physically inactive (OR 1.5; 95% CI, 1.1–2.0) and gynecologic cancer survivors were more likely to report current tobacco use (OR 1.8; 95% CI, 1.2–2.8).
Conclusions and Implications for Cancer Survivors Specific subgroups of cancer survivors are more likely to engage in unhealthy behaviors. Accurate assessment of who may derive the most benefit will aid public health programs to effectively target limited resources.

Keywords Cancer · Survivors · Life style · Behavior

A. Linsky (✉)
Section of General Internal Medicine, Department of Medicine,
Boston University School of Medicine,
801 Massachusetts Avenue, 2nd Floor, GIM,
Boston, MA 02118, USA
e-mail: Amy.Linsky@bmc.org

J. Nyambose
Massachusetts Comprehensive Cancer Prevention and Control Program, Bureau of Community Health and Prevention,
Massachusetts Department of Public Health,
Boston, MA, USA

T. A. Battaglia
Women's Health Unit, Section of General Internal Medicine and Women's Health Interdisciplinary Research Center,
Evans Department of Medicine,
Boston University School of Medicine,
Boston, MA, USA

Background

With continued improvements in detection and treatment, more Americans are surviving a diagnosis of cancer. Defining a survivor as someone with any history of a cancer diagnosis, regardless of time since diagnosis, there were nearly 12 million American survivors in 2007 [1]. These survivors are at risk for recurrence, secondary cancers, and other medical problems, including cardiovascular disease and diabetes [2].

Risk for subsequent health problems may result from cancer treatments, genetics, or lifestyle behaviors [3]. Estimates indicate that one third of cancer deaths are related to tobacco and another third are due to physical inactivity and dietary habits [4]. While a cancer diagnosis is conceptualized as a "teachable moment," behavior change made post-diagnosis is often not maintained [3]. As such, tertiary prevention via adoption and maintenance of healthy behaviors and avoidance of unhealthy habits has potential

to reduce adverse health consequences and improve quality of life for survivors [3, 5].

Studies of national survey data suggest that, as a group, cancer survivors have similar lifestyles as those without a cancer history [6–8]; however, certain survivors are at increased risk for unhealthy behaviors. Younger survivors are more often current smokers [6, 7], yet concurrently are more physically active [6]. Meanwhile, prostate cancer survivors have a higher prevalence of moderate-to-heavy alcohol use [8].

Prior studies have conflicting findings regarding unhealthy behaviors in cancer survivors, in part due to differences in methodology and unmeasured confounders, such as physical and mental health status. Furthermore, regional variation may be obscured with national level analyses. Since many public health interventions are organized at the state level, detailing survivors' behaviors within an individual state may better direct limited resources and public health planning. Therefore, we sought to determine the prevalence of unhealthy lifestyle behaviors of Massachusetts cancer survivors by conducting a cross sectional analysis using data from the Behavioral Risk Factor Surveillance System (BRFSS). Because patient characteristics associated with specific cancer sites differ, we also assessed whether survivors of specific cancers were more or less likely to engage in unhealthy lifestyles.

Methods

Data source

The BRFSS is an annual state-based cross-sectional telephone survey of adults ≥18 years, established and coordinated by the Centers for Disease Control and Prevention (CDC) [9]. A range of information is collected, including health conditions, alcohol and drug use, sexual behaviors, and environmental exposures. Massachusetts added cancer survivorship questions to the BRFSS Survey in 2006 to track behaviors of survivors.

Participants

There were 20,900 Massachusetts residents who were asked "Have you ever been diagnosed with any type of cancer?" in 2006–2008. The 18,197 respondents who reported no history of cancer comprise the non-cancer controls. The 2,339 subjects who responded affirmatively were then asked to name the cancer site (up to three). Individuals who refused or did not know cancer history or cancer site were excluded. We further removed those who did not report age and gender, reported non-melanoma skin cancer only, or had inconsistent gender/cancer combinations, yielding 1,670 cancer survivors in our study.

Cancer survivors were further classified by primary site of cancer. We focused specifically on the sites with the highest frequencies of responses: female breast, gynecologic, prostate, and colorectal (CRC). Gynecologic survivors were those who answered that they were diagnosed as having cervical, uterine, or ovarian cancer—a single combined answer option in the BRFSS. Respondents with other cancer sites or a history of multiple cancers formed separate subgroups of "Single—other" and "Multiple," respectively.

Demographics and covariates

Self-reported information was obtained for age, race, ethnicity, education, employment status, marital status, health insurance, and physical and mental health status. Race and ethnicity were recoded into white non-Hispanic vs. other. We dichotomized education, marital status, and health insurance. Employment status had three levels (employed, unemployed, retired).

Physical health was assessed using the question "For how many days during the past 30 days was your physical health not good?" We categorized responses into ≥14 day vs. <14 days, which corresponds with physical activity levels [10]. Mental health was assessed by asking "For how many days during the past 30 days was your mental health not good?" Frequent Mental Distress was considered present if the subject reported ≥14 days, consistent with CDC standards and other studies [11, 12].

Behavioral outcomes

Tobacco use Smoking behavior was measured using combinations of current and former cigarette use. "Current smokers" reported ≥100 lifetime cigarettes and current use. "Former smokers" reported ≥100 lifetime cigarettes without current use. Subjects with <100 lifetime cigarettes were classified as "Never smokers." All analyses assessing former tobacco use excluded never smokers since they could not become either a former or current smoker.

Alcohol use Heavy alcohol use was defined as an average intake in the past 30 days of 60 drinks for a man and 30 drinks for a woman, consistent with general recommendations in the absence of universal guidelines for alcohol consumption [4].

Physical inactivity Respondents were asked "During the past month...did you participate in any physical activities such as running, calisthenics, golf, gardening or walking for

exercise?" If they answered negatively, they were designated physically inactive.

Statistical analyses

To account for the complex sample survey design, all analyses were weighted to reflect the probability of selection of a telephone number, the number of adults in a household, and differences in participation by gender and age to provide Massachusetts state-level estimates. All sample sizes reported are unweighted and all percentages are weighted. Due to the imbalance in opportunity for men and women to be included in the study population (two female cancers, one male cancer, one gender neutral cancer) all analyses were stratified by gender. Within this stratification, analyses were conducted on survivors as a group and by subgroups of specific cancer sites.

We used descriptive statistics and chi-square tests to examine demographics and differences in categorical baseline characteristics. Differences in age were tested with *t*-tests. We assessed for colinearity of characteristics with non-parametric methods.

We determined both unadjusted prevalence and then age adjusted prevalence of each behavioral outcome. Adjusted rates were calculated using the direct method to the year 2000 Census Massachusetts population for each behavioral outcome for cancer survivors overall and site-specific survivors. Age adjusted logistic regression models determined the odds of each outcome for survivors and site-specific survivors compared to controls. We then modeled each behavior adjusting for age, race/ethnicity, education, and physical and mental health status, as these factors have potential to impact the adoption of health behaviors. Missing data were excluded from analysis.

Statistical significance was set at alpha=0.05. All analyses were performed using version 9.2 of SAS software (SAS Institute Inc, Cary, North Carolina). This study was approved by the Institutional Review Boards of the MA Department of Public Health and Boston University Medical Center.

Results

Population characteristics

Cancer survivors

Among the unweighted study sample of 19,867 respondents, there were more women than men (64% vs. 36%). In gender stratified samples, the weighted prevalence of having any cancer history was 4.9% for men and 7.7% for women (Table 1). Compared to same gender controls, survivors of any cancer were older, more likely to be white non-Hispanic, not be currently employed, have health insurance and report more poor physical health days (Table 2). There were no differences in poor mental health days for men or women. Among women, cancer survivors were less likely to have any college education.

Site-specific survivors (data not shown)

Among men, prostate cancer survivors (n=244) and CRC survivors (n=52) were older than controls (mean age 72 vs. 66 vs. 44 years, respectively). They were more likely to be white non-Hispanic, not be currently employed, and report more poor physical health days.

Among women, breast cancer survivors (n=479) and CRC survivors (n=72) were oldest (mean age 65 and 70 years, respectively) compared to gynecologic cancer survivors (n=275) and controls (51 and 46 years, respectively). Gynecologic survivors had the lowest prevalence of any college education. Compared to the breast and CRC survivors, they had the highest prevalence of ≥14 days poor physical health. Although not statistically significant, gynecologic survivors were more likely to have Frequent Mental Distress (14%).

Table 1 Gender stratified Massachusetts cancer survivor distribution, 2006–2008

	Men (n=6,709) n (%±SE)	Women (n=11,488) n (%±SE)
No history of cancer	6,709 (95.1±0.3)	11,488 (92.3±0.3)
Any cancer	516 (4.9±0.3)	1,154 (7.7±0.3)
Specific cancers		
Colorectal	52 (0.5±0.1)	72 (0.5±0.1)
Prostate	244 (2.2±0.2)	–
Breast	–	479 (3.1±0.2)
Gynecologic[a]	–	275 (2.1±0.2)
Single other site	188 (1.9±0.2)	264 (1.8±0.2)
Multiple sites	32 (0.3±0.1)	64 (0.3±0.1)

All percentages are weighted. Totals may not equal 100% due to rounding.

Data source: MA Behavioral Risk Factor Surveillance System, 2006–2008.

Abbreviations: *SE* standard error

[a] Gynecologic cancer includes cervical, uterine, and ovarian.

Table 2 Baseline characteristics by cancer survivorship and gender, 2006–2008

Characteristic	Men			Women		
	Non-cancer controls (*n*=6709) % (SE)	Survivors (*n*=516) % (SE)	*p*-value	Non-cancer controls (*n*=11,488) % (SE)	Survivors (*n*=1,154) % (SE)	*p*-value
Age, mean (95% CI)	44 (43–45)	67 (65–68)	<0.001	46 (46–47)	61 (59–64)	<0.001
Race/Ethnicity			0.004			<0.001
Non-Hispanic White	81 (0.8)	90 (2.3)		83.5 (0.6)	94 (0.9)	
Other	19 (0.8)	10 (2.3)		16.5 (0.6)	6 (0.9)	
Missing, n	70	5		76	10	
Education			0.42			0.001
College or more	68 (1.0)	65 (3.0)		68.3 (0.7)	61 (2.2)	
High School or less	32 (1.0)	35 (3.0)		31.7 (0.7)	39 (2.2)	
Missing, n	14	3		16	1	
Employment			<0.001			<0.001
Employed	74 (0.8)	32 (3.0)		61 (0.7)	40 (2.2)	
Unemployed	14 (0.8)	15 (2.3)		25 (0.7)	25 (2.0)	
Retired	12 (0.5)	53 (3.1)		14 (0.4)	36 (2.0)	
Missing, n	16	0		14	2	
Married			<0.001			<0.001
Yes	65 (1.0)	75 (2.5)		61 (0.8)	54 (2.2)	
No	35 (1.0)	25 (2.5)		39 (0.8)	46 (2.2)	
Missing, n	25	1		39	2	
Health Insurance			<0.001			0.017
Yes	90 (0.7)	99 (0.3)		95 (0.4)	97 (0.7)	
No	10 (0.7)	1 (0.3)		5 (0.4)	3 (0.7)	
Missing, n	20	0		18	2	
Physical Health			<0.001			<0.001
≥14 unhealthy days	8 (0.5)	20 (2.5)		10 (0.4)	19 (1.7)	
<14 unhealthy days	92 (0.5)	80 (2.5)		90 (0.4)	81 (1.7)	
Missing, n	77	15		200	42	
Mental Health			0.43			0.26
≥14 unhealthy days[a]	8 (0.5)	9 (1.8)		10.5 (0.5)	12 (1.5)	
<14 unhealthy days	92 (0.5)	91 (1.8)		89.5 (0.5)	88 (1.5)	
Missing, n	89	7		164	17	

All percentages are weighted. Totals may not equal 100% due to rounding.

MA Behavioral Risk Factor Surveillance System, 2006–2008.

[a] Frequent Mental Distress is ≥14 unhealthy days of Mental Health.

Abbreviations: *SE* standard error

Prevalence of health behaviors

Cancer survivors

There was no difference in the age adjusted prevalence of heavy alcohol use or physical inactivity for male survivors compared to controls, but survivors were more likely to report former smoking (45% vs. 29%) and less likely to report never smoking (31% vs. 53%) (Table 3). Likewise, female cancer survivors were similar to controls in heavy alcohol use and physical inactivity, but they were more likely to report current smoking (24% vs. 16%) or former smoking (39% vs. 25%) and less likely to report never smoking (42% vs. 60%).

Site-specific survivors (data not shown)

Small sample size for site-specific male survivors precluded age adjusting, but in unadjusted analyses, men with a history of prostate cancer or CRC had comparable heavy alcohol use and physical inactivity as controls. In age adusted analyses of women survivors of breast and

Table 3 Age adjusted prevalence of behavioral risk factors by cancer survivorship and gender, 2006–2008

Behavior	Men		Women	
	Non-cancer controls (*n*=6,709) % (SE)	Survivors (*n*=516) % (SE)	Non-cancer controls (*n*=11,488) % (SE)	Survivors (*n*=1,154) % (SE)
Heavy drinking				
Yes	6.3 (0.5)	8.4 (4.5)	5.0 (0.3)	9.3 (2.4)
No	93.7 (0.5)	91.6 (4.5)	95.0 (0.3)	90.7 (2.4)
Missing, n	170	11	234	17
Smoking				
Current	18.5 (0.8)	24.0 (6.2)	15.5 (0.6)	24.0 (3.2)
Former	29.0 (0.7)	45.1 (6.4)	25.0 (0.6)	33.6 (3.1)
Never	52.5 (0.9)	30.9 (5.8)	59.5 (0.7)	42.4 (3.6)
Missing, n	30	0	57	6
Physical inactivity				
Inactive	19.9 (0.8)	31.4 (6.4)	22.5 (0.6)	27.5 (2.9)
Active	80.1 (0.8)	68.6 (6.4)	77.5 (0.6)	72.5 (2.9)
Missing, n	3	0	2	1

All percentages are weighted. Totals may not equal 100% due to rounding. Age adjusted by the direct method to the year 2000 Census Massachusetts population using the age groups 18–49 years, 50–59 years, 60–69 years, and 70 years or older. MA Behavioral Risk Factor Surveillance System, 2006–2008.

Abbreviations: *SE* standard error

gynecologic cancers, there were no differences from controls in heavy alcohol use or physical inactivity. Gynecologic survivors had more current smoking than controls (32% vs. 16%) and breast cancer survivors reported more former smoking than controls (43% vs. 25%).

Adjusted odds of health behaviors

Cancer survivors

In multivariable models, male survivors were no more likely to have heavy alcohol use, current or former tobacco use, or be physically inactive than controls (Table 4). Female survivors had similar alcohol and tobacco use as controls, but had greater odds of being physically inactive [Odds Ratio (OR) 1.5; 95% CI, 1.2–1.8].

Site-specific survivors

Separate multivariable models for male survivors of CRC or prostate cancer found no differences from controls for any unhealthy behavior (Table 4). Models for female CRC survivors also did not detect any statistically significant differences from controls. However, breast cancer survivors continued to have greater physically inactivity (OR 1.5; 95% CI, 1.1–2.0), while gynecologic cancer survivors had higher odds of current smoking (OR 1.8; 95% CI, 1.2–2.8).

Discussion

We illustrate unhealthy behaviors of cancer survivors residing in Massachusetts. Specifically, after adjusting for physical and mental health, female survivors, especially of breast cancer, were more likely to be physically inactive. Gynecologic cancer survivors were more likely to currently smoke, even after controlling for physical and mental health status. Given the known benefits of tobacco cessation and physical activity, these results highlight specific survivor subgroups to target interventions aimed at reducing health risks.

An elevated rate of former smoking among survivors overall is similar to others' findings and may reflect successful secondary prevention [7, 13]. Concurrently, the higher odds of current smoking in gynecologic cancer survivors (which includes cervical cancer) is consistent with previously reported cervical cancer survivors' behavior [6–8] and warrants further attention. We expanded on this association by controlling for mental health since higher rates of depression are seen in cervical cancer survivors [14], and depression is associated with tobacco use [15]. We also adjusted for race to account for the higher proportion of poor health behaviors in minority survivors compared to white survivors [16]. Finally, our findings may be related to the fact that tobacco increases host susceptibility to human papilloma virus [17].

Table 4 Adjusted odds of health behaviors

	Heavy drinking OR (95% CI)	Current smoking OR (95% CI)	Former smoking OR (95% CI)	Physical inactivity OR (95% CI)
Male non-cancer control (ref)	1.0	1.0	1.0	1.0
Male survivors, any cancer	1.3 (0.7–2.3)	0.8 (0.5–1.3)	1.3 (0.8–2.1)	1.0 (0.7–1.4)
Male non-cancer control (ref)	1.0	1.0	1.0	1.0
Male survivors, CRC	1.7 (0.5–5.7)	0.5 (0.1–1.9)	3.1 (0.7–13.5)	0.7 (0.3–1.6)
Male non-cancer control (ref)	1.0	1.0	1.0	1.0
Male survivors, prostate	1.1 (0.5–2.2)	0.8 (0.4–1.6)	0.9 (0.4–1.8)	0.8 (0.5–1.2)
Female non-cancer control (ref)	1.0	1.0	1.0	1.0
Female survivors, any cancer	1.4 (0.9–2.2)	1.1 (0.8–1.5)	1.2 (0.9–1.6)	1.5 (1.2–1.8)
Female non-cancer control (ref)	1.0	1.0	1.0	1.0
Female survivors, CRC	0.7 (0.1–3.5)	1.0 (0.3–3.8)	0.8 (0.2–3.3)	1.4 (0.6–3.6)
Female non-cancer control (ref)	1.0	1.0	1.0	1.0
Female survivors, breast	1.2 (0.6–2.2)	0.8 (0.5–1.3)	1.5 (0.9–2.4)	1.5 (1.1–2.0)
Female non-cancer control (ref)	1.0	1.0	1.0	1.0
Female survivors, Gyn[a]	1.9 (0.9–4.3)	1.8 (1.2–2.8)	0.8 (0.4–1.5)	1.3 (0.8–2.1)

Multivariable logistic regression adjusted for age, race, education, and physical and mental health status.

Former smoking models only include former and current smokers (never smokers excluded).

MA Behavioral Risk Factor Surveillance System, 2006–2008.

Abbreviations: *CI* Confidence Interval, *CRC* Colorectal, *Gyn* Gynecologic.

[a] Gynecologic cancer includes cervical, uterine, and ovarian.

In evaluating physical inactivity, we used a relatively low threshold compared to other studies that required minimum exertion or duration. Even still, in adjusted analyses, female survivors were more likely to be physically inactive. When analyzed by cancer site, breast cancer survivors were significantly more inactive than controls. This contrasts to others' findings where after adjustment survivors were no different from controls [7], or even more likely to be active [8].

Different definitions of physical activity may explain these discrepancies or there may be uncontrolled confounding. Both poor physical health and depressive symptoms have been associated with lower physical activity in breast cancer survivors [18]. Grimmett et al. did adjust for the presence of arthritis [13], but other comorbidities have physical limitations. We controlled for self-perceived physical health, which has been associated with physical activity restriction [10].

Limitations

Our findings should be interpreted in the context of the following limitations. The BRFSS interviews non-institutionalized individuals with land telephones and results may not generalize to institutionalized populations or those with only cellular telephones. All cancer history is self-reported and subject to inaccuracy, but previous research has demonstrated high quality of cancer reporting via the BRFSS [19]. Data on cancer stage and treatments are unknown, but it is unclear how each would affect the impact a cancer diagnosis has on behavior change. Given the observational nature of the study we can only assess association, not causation. The small sample sizes for subgroups of specific cancers, especially colorectal and prostate, may have been underpowered to show an association. However, this does not detract from those results which were statistically significant. We are limited by lack of data on temporality, including pre-diagnosis behaviors and time since diagnosis. Behaviors pre-diagnosis can predict post-diagnosis habits [5], and we cannot determine behavioral changes due to a cancer diagnosis. Additionally, behavioral changes may occur proximal to the diagnosis, but not be sustained with time [18]. Finally, the BRFSS options for cancer site combined cervical, uterine, and ovarian cancers into one response choice of gynecologic cancer, precluding further cancer type stratification. The relatively younger age of the gynecologic cancer survivors, along with the presence of a screening test and better survival, may indicate a higher proportion are cervical cancer survivors, but this needs to be addressed in future studies.

Implications for cancer survivors

This study provides guidance to clinicians and public health professionals for targeted interventions to improve

healthy behaviors in cancer survivors. Recognizing patterns on the state level enables the appropriate public health agencies to direct funds and programming resources to those most in need. This is highlighted by the fact that MA has similar, but not identical, findings of cancer incidence and health behaviors as compared to a national sample [20]. Age adjusted invasive cancer incidence and mortality in MA is greater than the national rate. Further, MA state level estimates of heavy drinking for men and women are higher than national prevalence estimates, while smoking rates and physical inactivity are lower. As individual states may have different behavior profiles for survivors, programs can be suitably tailored. Perhaps most striking in our study is the high rate of current tobacco use among gynecologic survivors. Unlike age, genetics, and other immutable factors, smoking is modifiable. Its numerous associations with risk for cancer and other medical conditions make it a prime target of intervention. This subgroup is younger and investments in their health may have more time to show health and economic returns.

Greater physical inactivity in breast cancer survivors also raises concern for subsequent adverse health. Vigorous exercise is associated with lower all-cause mortality in cancer survivors [21]. Engagement in physical activity has been shown to benefit breast cancer survivors, improving health outcomes and health related quality of life [22, 23]. Even with a liberal definition of physical activity, we found that breast cancer survivors had more inactivity. These survivors, who continue to grow in number, may need additional resources targeted to increasing their activity levels.

Welcomed improvements in diagnosis and treatment will also lead to a larger population of cancer survivors, with attendant increased risk for other cancers and medical conditions. Successful secondary and tertiary prevention to promote tobacco cessation and adoption of physical activity may mitigate some of these risks. With accurate assessment of populations who may derive the most benefit from interventions, public health programs can most effectively direct limited resources. Ongoing research is warranted to determine the most effective means to initiate and maintain adoption of healthy lifestyle behaviors.

Acknowledgements Thank you to Michael Winter, MPH for his statistical assistance.

References

1. SEER Cancer Statistics Review, 1975–2007. National Cancer Institute, 2010. Available at: http://seer.cancer.gov/csr/1975_2007/. Accessed July 9, 2010.
2. Gallicchio L, Kalesan B, Hoffman SC, Helzlsouer KJ. Non-cancer adverse health conditions and perceived health and function among cancer survivors participating in a community-based cohort study in Washington County, Maryland. J Cancer Surviv. 2008;2:12–9.
3. Demark-Wahnefried W, Pinto BM, Gritz ER. Promoting health and physical function among cancer survivors: potential for prevention and questions that remain. J Clin Oncol. 2006;24:5125–31.
4. American Cancer Society. Cancer Facts & Figures 2008. Atlanta: American Cancer Society; 2008.
5. Demark-Wahnefried W, Jones LW. Promoting a healthy lifestyle among cancer survivors. Hematol Oncol Clin North Am. 2008;22:319–42. viii.
6. Coups EJ, Ostroff JS. A population-based estimate of the prevalence of behavioral risk factors among adult cancer survivors and noncancer controls. Prev Med. 2005;40:702–11.
7. Eakin EG, Youlden DR, Baade PD, et al. Health behaviors of cancer survivors: data from an Australian population-based survey. Cancer Causes Control. 2007;18:881–94.
8. Bellizzi KM, Rowland JH, Jeffery DD, McNeel T. Health behaviors of cancer survivors: examining opportunities for cancer control intervention. J Clin Oncol. 2005;23:8884–93.
9. Behavioral Risk Factor Surveillance System: Frequently Asked Questions. Available at: http://www.cdc.gov/brfss/faqs.htm. Accessed July 6, 2010.
10. Brown DW, Balluz LS, Heath GW, et al. Associations between recommended levels of physical activity and health-related quality of life. Findings from the 2001 Behavioral Risk Factor Surveillance System (BRFSS) survey. Prev Med. 2003;37:520–8.
11. Centers for Disease Control and Prevention. Measuring Healthy Days: Population Assessment of Health-Related Quality of Life. Available at: http://www.cdc.gov/hrqol/pdfs/mhd.pdf. Accessed July 23, 2010.
12. Hennessy CH, Moriarty DG, Zack MM, Scherr PA, Brackbill R. Measuring health-related quality of life for public health surveillance. Public Health Rep. 1994;109:665–72.
13. Grimmett C, Wardle J, Steptoe A. Health behaviours in older cancer survivors in the English Longitudinal Study of Ageing. Eur J Cancer. 2009;45:2180–6.
14. McCorkle R, Tang ST, Greenwald H, Holcombe G, Lavery M. Factors related to depressive symptoms among long-term survivors of cervical cancer. Health Care Women Int. 2006;27:45–58.
15. Lasser K, Boyd JW, Woolhandler S, Himmelstein DU, McCormick D, Bor DH. Smoking and mental illness: a population-based prevalence study. JAMA. 2000;284:2606–10.
16. Schootman M, Deshpande AD, Pruitt SL, Aft R, Jeffe DB. National estimates of racial disparities in health status and behavioral risk factors among long-term cancer survivors and non-cancer controls. Cancer Causes Control. 2010;21(9):1387–95.
17. Hildesheim A, Herrero R, Castle PE, et al. HPV co-factors related to the development of cervical cancer: results from a population-based study in Costa Rica. Br J Cancer. 2001;84:1219–26.
18. Emery CF, Yang HC, Frierson GM, Peterson LJ, Suh S. Determinants of physical activity among women treated for breast cancer in a 5-year longitudinal follow-up investigation. Psychooncology. 2009;18:377–86.
19. Kapp JM, Jackson-Thompson J, Petroski GF, Schootman M. Reliability of health-related quality-of-life indicators in cancer survivors from a population-based sample, 2005, BRFSS. Public Health. 2009;123:321–5.

20. Centers for Disease Control and Prevention. National Center for Chronic Disease Prevention and Health Promotion Chronic Disease Indicators. Available at: http://apps.nccd.cdc.gov/cdi/Default.aspx. Accessed October 26, 2010.
21. Hamer M, Stamatakis E, Saxton JM. The impact of physical activity on all-cause mortality in men and women after a cancer diagnosis. Cancer Causes Control. 2009;20:225–31.
22. Holmes MD, Chen WY, Feskanich D, Kroenke CH, Colditz GA. Physical activity and survival after breast cancer diagnosis. JAMA. 2005;293:2479–86.
23. Alfano CM, Smith AW, Irwin ML, et al. Physical activity, long-term symptoms, and physical health-related quality of life among breast cancer survivors: a prospective analysis. J Cancer Surviv. 2007;1:116–28.

Patient-Provider Language Concordance and Colorectal Cancer Screening

Amy Linsky, MD, MSc[1], Nathalie McIntosh, MSc[2], Howard Cabral, PhD[3], and Lewis E. Kazis, ScD[4]

[1]General Internal Medicine, Boston Medical Center, Boston, MA, USA; [2]Department of Health Policy and Management, Boston University School of Public Health, Boston, MA, USA; [3]Biostatistics, Boston University, Boston, MA, USA; [4]Center for the Assessment of Pharmaceutical Practices (CAPP)/Department of Health Policy and Management, Boston University School of Public Health, Boston, MA, USA.

BACKGROUND AND OBJECTIVE: Patient-provider language barriers may play a role in health-care disparities, including obtaining colorectal cancer (CRC) screening. Professional interpreters and language-concordant providers may mitigate these disparities.

DESIGN, SUBJECTS, AND MAIN MEASURES: We performed a retrospective cohort study of individuals age 50 years and older who were categorized as English-Concordant (spoke English at home, n=21,594); Other Language-Concordant (did not speak English at home but someone at their provider's office spoke their language, n=1,463); or Other Language-Discordant (did not speak English at home and no one at their provider's spoke their language, n=240). Multivariate logistic regression assessed the association of language concordance with colorectal cancer screening.

KEY RESULTS: Compared to English speakers, non-English speakers had lower use of colorectal cancer screening (30.7% vs 50.8%; OR, 0.63; 95% CI, 0.51–0.76). Compared to the English-Concordant group, the Language-Discordant group had similar screening (adjusted OR, 0.84; 95% CI, 0.58–1.21), while the Language-Concordant group had lower screening (adjusted OR, 0.57; 95% CI, 0.46–0.71).

CONCLUSIONS: Rates of CRC screening are lower in individuals who do not speak English at home compared to those who do. However, the Language-Discordant cohort had similar rates to those with English concordance, while the Language-Concordant cohort had lower rates of CRC screening. This may be due to unmeasured differences among the cohorts in patient, provider, and health care system characteristics. These results suggest that providers should especially promote the importance of CRC screening to non-English speaking patients, but that language barriers do not fully account for CRC screening rate disparities in these populations.

KEY WORDS: language concordance; cancer screening; disparities.
J Gen Intern Med 26(2):142–7
DOI: 10.1007/s11606-010-1512-9

Received March 19, 2010
Revised August 17, 2010
Accepted August 18, 2010
Published online September 21, 2010

BACKGROUND

The United States has tremendous ethnic and linguistic diversity. According to the 2005–2007 American Community Survey, minorities comprise 26% of the population, and nearly 20% of Americans speak a language other than English at home. By 2050, it is projected that minorities will make up about half of the US population, with a similar increase in individuals speaking a language other than English at home.[1] Compared to white non-Hispanics, minorities use fewer preventive services, including colorectal cancer (CRC) screening.[2] Language plays a role in these health-care disparities.[3, 4] Language barriers may undermine medical communication, lead to inaccurate diagnosis, and contribute to poorer management or treatment adherence.[5]

Of individuals who do not speak English at home, roughly 44% speak English "less than well."[1] Patient-provider communication problems may be common for individuals who have limited English proficiency (LEP).[1,2,6] Compared to those with English proficiency, LEP patients are more likely to have difficulty understanding medical explanations,[7] getting information,[8] and have worse management of care.[9] LEP patients are less likely to have preventive[10] or primary care services,[11] access to care,[12] or be satisfied with provider communication.[13] Access to, and the quality of, care for LEP patients can be improved by using professional interpreters or language-concordant providers.[4]

Colorectal cancer screening is recommended in the routine care of older patients in the US and may be compromised because of language barriers. CRC is the third most prevalent cancer in the US.[2] Although many minorities have lower rates of CRC compared to white non-Hispanics, they tend to be diagnosed at a later stage of disease and have higher mortality rates.[2] The US Preventive Services Task Force recommends colonoscopy every 10 years in adults aged 50–75 years. Other recommendations include flexible sigmoidoscopy every 5 years or home-based fecal occult blood test (FOBT) every year.[14] It is estimated that 60% of CRC deaths could be prevented if all persons age 50 years and older were screened.[15]

Current rates of CRC screening are less than 60%, with lower rates in minorities.[16] Language appears to be a significant factor.[17] Compared to white non-Hispanics, Spanish-speaking Hispanics were 43% less likely to receive CRC screening.[18] Communication problems when discussing cancer screening are also documented with Vietnamese-Americans.[19] Furthermore, there is evidence that fewer

providers discuss CRC screening with non-English speaking patients[20] even when translators are available.[21]

OBJECTIVE

Lack of physician recommendation is often the primary reason patients are not current with guidelines.[19] Discussion of screening for CRC is complicated and time consuming, and may be omitted or abbreviated when there are language barriers.[22] A recent study showed that patients who spoke Spanish at home were less likely to receive CRC screening compared to patients who spoke English at home, even after controlling for English proficiency and patient characteristics.[23] However, that study did not take into account whether someone at the provider's office spoke the patient's preferred language. The purpose of our study is to assess the association of language concordance with CRC screening rates in patients who do not speak English at home compared to rates in those who do.

DESIGN AND SUBJECTS

We analyzed data from the Medical Expenditures Panel Survey (MEPS), a nationally representative survey of non-institutionalized US civilians with 2 years of longitudinal follow-up. We used the Consolidated Household and Medical Conditions files from the Household Component Survey and the Self-Administered Questionnaire. We merged 2002, 2004 and 2006 data, choosing alternate years to ensure distinct respondents, creating a sample of 107,720 subjects. Individuals with a self-reported history of colon or rectal cancer (*International Classification of Disease 9-CM* codes 153, 154) and those less than 50 years were excluded, as were individuals who did not have complete responses for all variables of interest. Individuals greater than 75 years were included given lack of consensus on an upper age limit for screening. Our final study sample was 23,297 subjects, representing 222 million individuals.

MAIN MEASURES

To create cohorts of patient-provider language concordance, we combined responses to the questions "What language is spoken in your home most of the time?" and "Does someone at your provider's speak the language you prefer or provide translator services?" If English was spoken at home, the subject was categorized as *English-Concordant*. If English was not spoken at home and someone at the provider's spoke the respondent's preferred language or offered translation services, the subject was categorized as *Other Language-Concordant*. Subjects who reported not speaking English at home and denied that someone at their provider's spoke their preferred language or offered translation services formed the third cohort, *Other Language-Discordant* (Fig. 1).

We assessed CRC screening using self-reported rates of FOBT and endoscopy. Given that patients may not have FOBT exactly within 12 months, we considered tests performed within 2 years prior to the date of MEPS survey completion to be current with recommendations. In MEPS data, responses for sigmoidoscopy and colonoscopy were combined into a single variable with time frame choices either within or greater than 5 years of the survey date. Therefore, if an individual had FOBT within 2 years or endoscopy within 5 years, they were classified as being current with CRC screening.

Covariates included: self-reported race/ethnicity, age, education, marital status, family income, employment status, time since last check-up, and health insurance status. Race and ethnicity were combined into a variable with five categories: white non-Hispanic, black, Hispanic, Asian, and other. Age had three categories: age 50–<65 years, 65–<75 years, and age 75–85 years (this category may include subjects >85 years as

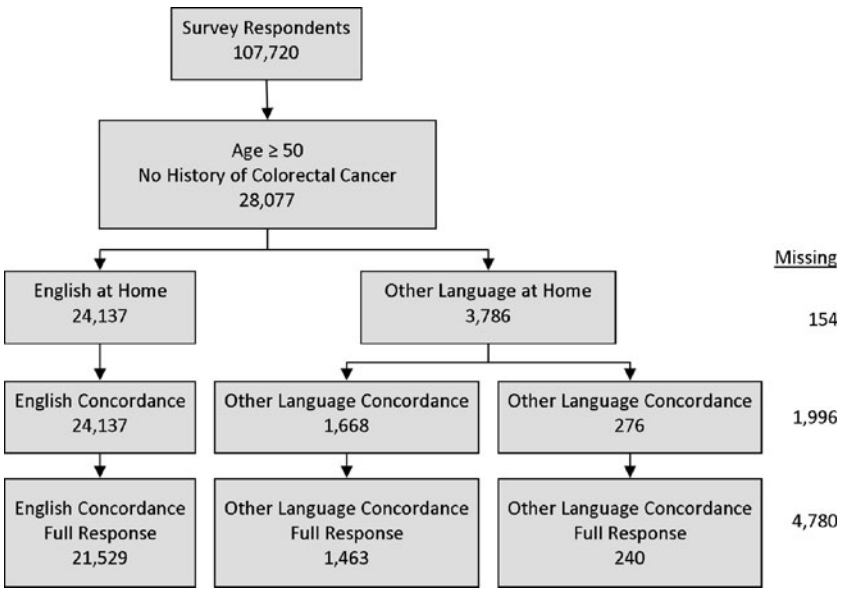


Figure 1. Sample cohorts.

MEPS top coded ages at 85). Education had four categories: no degree, high school or equivalent, some college or greater, or other. Marital status had two categories: married or not married. Total family income had four categories defined by the federal poverty line (FPL): poor/near-poor (<125% FPL), low income (125–<200% FPL), middle income (200–<400% FPL), and high income (≥400% FPL). Employment status had two categories: employed or unemployed. Time since the last checkup had three categories: within 2 years, greater than 2 years, and never. Health insurance status had three categories: any private insurance, public insurance only, and no insurance.

Co-morbidities are often considered when recommending CRC screening; however, these indices are not included in MEPS. As a proxy, the Physical Component Score (PCS) and Mental Component Score (MCS) of the 12 Item Short Form Health Survey (SF-12) were used. These scores have been well validated and are standardized with a mean of 50 for the general population.[24] Scores were converted into two categories: scores less than the sample median ("low" PCS/MCS) and those greater than or equal to the median ("high" PCS/MCS). Sample PCS and MCS medians were 47 and 54, respectively. Of note, Medicare coverage of endoscopy for average-risk adults began in 2001, so year of survey was also included. We included US region to account for regional variations. The following provider-level variables were also included: provider race, ethnicity, and sex; and provider type and specialty.

Analysis

To account for the complex sample design, survey statistical procedures were used. Weighted prevalence and standard error (SE) estimates were calculated for independent variables using MEPS survey weights, and χ^2 tests assessed for differences between cohorts. Variables with proportion of missing responses greater than 65% (provider characteristics) were eliminated. For the remaining variables, individuals with complete data were compared to those with missing data to assess generalizability. All subsequent analyses were done on samples with complete data for all variables retained (n= 23,297).

Bivariate odds ratios (ORs) and 95% confidence intervals (CIs) were used to evaluate the associations between each independent variable and receipt of CRC screening. Variables were independently assessed for confounding or effect modification of language concordance. Those yielding a ≥10% change in the magnitude of effect were considered as potential confounders.

We used multivariate logistic regression to determine the association of language concordance with CRC screening. We included in the model those variables determined *a priori* to be potentially associated with screening (sex, age, time since last checkup, marital status, employment status, year, region), as well as variables found to be confounders (race/ethnicity, education, family income, and health insurance status) or effect modifiers (PCS).

Sensitivity Analyses

We evaluated our definitions of both the primary explanatory variable and outcome variables. Patient language was re-defined by comfort level speaking English using the question "Are you comfortable conversing in English?" Individuals with LEP (those who responded "No") were then grouped by whether someone at their provider's spoke their preferred language or offered translation services. In this way, three cohorts based on English proficiency were created: English-proficient, LEP-Concordant and LEP-Discordant. A simple logistic regression using these re-defined cohorts was compared to that using the original cohorts based on language spoken at home. In addition, receipt of CRC screening was re-defined in several ways. We assessed receipt of FOBT alone, endoscopy alone, endoscopy ever, and any screen ever.

All prevalence, odds ratio, and variance estimates are presented from weighted analyses unless otherwise specified. Statistical significance was set at α=0.05. All analyses were conducted with SAS (version 9.2, SAS Institute Inc., Cary, NC). This study was granted exempt status by the Boston University Institutional Review Board.

KEY RESULTS

The final study sample of 23,297 represents 222 million individuals age 50 years or older with no history of CRC. The vast majority of respondents spoke English at home (96%). Overall, most were white non-Hispanic (81.1%), had at least a high school education (75.5%), were married (61.8%), were aged 50–64 years (57.8%), were female (54.7%), and were employed (51.3%). Few were uninsured (6.1%).

The English-Concordant cohort was predominantly white non-Hispanic (83.7%), more likely to have a high school education, and to have high income and private insurance (Table 1). The Other Language-Concordant cohort had the highest prevalence of Hispanics (60.9%), was the least educated, and most likely to be poor, have public health insurance, and be from the west. The Other Language-Discordant cohort had the highest percentage of Asians (29.4%) and those 75-85 years (29.6%), and was most likely to be unemployed or uninsured. The three cohorts were similar regarding sex, marital status, and time since last checkup. Given the large sample size, the cohorts were statistically different for all covariates except marital status (p=0.13).

The prevalence of CRC screening was greatest in the English-Concordant group, followed by the Other Language-Discordant group, and then the Other Language-Concordant group (50.8% vs 37.9% vs 28.9%, respectively). Compared to the English-Concordant cohort, the unadjusted odds of being current with CRC screening for Other Language-Concordant patients was 0.40 (95% CI, 0.33–0.47) and for Other Language-Discordant patients was 0.59 (95% CI, 0.42–0.84) (Table 2).

After adjusting for confounding, demographic and socioeconomic variables, the odds of being current with CRC screening for those who did not speak English at home was lower compared to those who did (30.7% vs 50.8%, respectively; OR, 0.63; 95% CI, 0.51–0.76).

When looking at patient-provider language concordance determined by language spoken at home and if someone at the provider's spoke the patient's preferred language, the adjusted odds of being current with CRC screening was lower for those in the Other Language-Concordant cohort compared to those in the English-Concordant cohort (OR, 0.57; 95% CI,

Table 1. Baseline Characteristics of Study Population by Language Concordance

Variable	English concordance (N=21,594)		Other language concordance (N=1,463)		Other language discordance (N=240)	
	n	Weighted % (SE)	n	Weighted % (SE)	n	Weighted % (SE)
Male sex	9,402	45 (0.25)	559	43 (1.32)	86	39 (3.10)
Age						
50–64	12,420	58.0 (0.58)	802	54.5 (2.19)	125	48.1 (4.15)
65–74	4,896	22.1 (0.41)	355	25.8 (1.65)	54	22.3 (3.43)
75–85	4,278	19.9 (0.47)	306	19.6 (1.52)	61	29.6 (3.93)
Race						
White non-Hispanic	16,350	83.7 (0.62)	102	14.0 (1.90)	41	23.2 (4.37)
Black	3,271	9.7 (0.48)	16	1.5 (0.57)	7	3.5 (1.42)
Hispanic	1,149	3.1 (0.22)	1,152	60.9 (2.65)	138	42.2 (4.88)
Asian	350	1.5 (0.15)	174	22.3 (2.56)	51	29.4 (4.30)
Other	474	2.0 (0.21)	19	1.4 (0.44)	3	1.7 (1.18)
Education						
No degree	4,387	15.5 (0.38)	1,058	63.3 (2.17)	144	49.5 (4.83)
High school Or GED	10,905	51.5 (0.55)	266	21.9 (1.65)	57	27.3 (4.03)
College or greater	4,801	25.5 (0.59)	107	11.8 (1.40)	32	19.2 (3.35)
Other	1,501	7.5 (0.24)	32	3.0 (0.62)	7	4.0 (1.59)
Married	12,754	61.7 (0.55)	915	65.9 (1.87)	144	61.5 (4.38)
Income*						
Poor/near poor	3,863	12.7 (0.36)	565	31.2 (1.73)	68	22.6 (3.63)
Low income	2,947	12.8 (0.35)	318	19.9 (1.54)	75	27.6 (4.13)
Middle income	5,951	27.6 (0.48)	410	31.8 (1.98)	49	22.6 (3.69)
High income	8,833	46.9 (0.67)	170	17.1 (1.67)	48	27.2 (4.32)
Insurance						
Any private	14,823	74.2 (0.54)	401	36.5 (2.46)	87	37.1 (4.10)
Public	5,281	20.0 (0.49)	820	50.8 (2.58)	111	48.5 (4.14)
Uninsured	1,490	5.8 (0.22)	242	12.7 (1.32)	42	14.4 (2.78)
Region						
Northeast	3,588	19.3 (0.86)	227	20.7 (2.20)	48	21.5 (3.82)
Midwest	5,033	23.9 (0.99)	56	5.9 (1.01)	23	12.9 (3.61)
South	8,661	37.1 (1.06)	545	29.9 (2.91)	88	30.9 (4.73)
West	4,312	19.7 (0.97)	835	43.6 (3.00)	81	34.7 (5.39)
Time since last checkup						
≤2 years	18430	85.9 (0.38)	1,269	88.0 (1.13)	215	90.2 (2.80)
>2 years	2468	11.2 (0.29)	107	7.1 (0.78)	10	3.1 (1.26)
Never	696	2.8 (0.19)	87	5.0 (0.80)	15	6.7 (2.44)
Employed	10,483	51.8 (0.55)	528	39.3 (2.10)	76	33.3 (3.71)
Physical component score†						
Low	10,625	46.0 (0.51)	874	57.2 (1.70)	146	56.6 (4.40)
High	10,969	54.0 (0.51)	589	42.8 (1.70)	94	43.4 (4.40)
Mental component score†						
Low	10,505	46.1 (0.44)	994	66.5 (1.80)	149	60.8 (3.63)
High	1,108	53.9 (0.44)	469	33.5 (1.80)	91	39.2 (3.63)

SE = standard error;
**Poor/near-poor (<125% Federal Poverty Level), low (125–<200 % FPL), middle (200–<400% FPL), high (≥400% FPL), †low = scores < median of the study population; high = scores ≥ median of the study population*

0.46–0.71). The Other Language-Discordant cohort did not statistically differ from the English-Concordant cohort (OR, 0.84; 95% CI, 0.58–1.21) (Table 2).

Sensitivity Analyses

Defining language concordance using English proficiency rather than language spoken at home did not change the patterns of association with CRC screening (LEP-Discordant: OR, 0.41; 95% CI, 0.20–0.83; LEP-Concordant: OR, 0.27; 95% CI, 0.19–0.37, referent to English-proficient). Furthermore, using different definitions of CRC screening (e.g., FOBT only, endoscopy only, endoscopy ever, and any screen ever) also yielded similar patterns of association, with lower rates of CRC screening in individuals who did not speak English at home compared to those who did, and higher rates of CRC screening in the Other Language-Discordant cohort compared to the Other Language-Concordant cohort (data not shown).

CONCLUSIONS

We found that individuals who do not speak English at home are less likely to be adherent with CRC screening compared to those who do. This is consistent with other reports.[3,10,23] However, in our adjusted model, we found that the Other Language-Discordant cohort is as likely as the English-Concordant cohort to be adherent to CRC screening guidelines, while the Other Language-Concordant cohort is less likely to be adherent. This was unexpected. We hypothesized that the Other Language-Concordant cohort would experience better

Table 2. Association of Independent Variables with Colorectal Cancer Screening

Variable (n in thousands)	Unadjusted odds ratio* (95% Confidence interval)	Adjusted odds ratio† (95% Confidence interval)
Concordance		
English (ref) (21.6)	1.0	1.0
Other concordance (1.4)	0.40 (0.33–0.47)	0.57 (0.46–0.71)
Other discordance (0.2)	0.59 (0.42–0.84)	0.84 (0.58–1.2)
Age		
50–64 (ref) (13.3)	1.0	1.0
65–74 (5.3)	1.74 (1.62–1.88)	1.55 (1.42–1.70)
75–85 (4.6)	1.39 (1.28–1.51)	1.22 (1.10–1.35)
Sex		
Male (ref) (10)	1.0	1.0
Female (13.3)	0.92 (0.87–0.97)	0.88 (0.83–0.94)
Race		
White non-Hispanic (ref) (16.5)	1.0	1.0
Black (3.3)	0.80 (0.72–0.89)	0.97 (0.88–1.08)
Hispanic (2.4)	0.57 (0.50–0.64)	0.92 (0.79–1.07)
Asian (0.6)	0.55 (0.45–0.68)	0.60 (0.48–0.76)
Other (0.5)	0.70 (0.57–0.86)	0.83 (0.67–1.03)
Income‡		
High income (ref) (9.1)	1.0	1.0
Middle income (6.4)	0.79 (0.73–0.86)	0.88 (0.81–0.96)
Low income (3.3)	0.67 (0.60–0.74)	0.74 (0.66–0.84)
Poor/near poor (4.5)	0.59 (0.53–0.65)	0.70 (0.62–0.80)
Education level		
College (ref) (4.9)	1.0	1.0
High school or GED (11.2)	0.72 (0.67–0.78)	0.74 (0.68–0.80)
No degree (5.6)	0.47 (0.43–0.51)	0.51 (0.46–0.58)
Other (1.5)	0.74 (0.65–0.84)	0.77 (0.67–0.89)
Insurance coverage		
Any private (ref) (15.3)	1.0	1.0
Public only (6.2)	0.83 (0.77–0.90)	0.87 (0.79–0.96)
Uninsured (1.7)	0.28 (0.24–0.32)	0.54 (0.46–0.63)
Year		
2002 (ref) (8.7)	1.0	1.0
2004 (7.1)	1.11 (1.03–1.19)	1.09 (1.01–1.18)
2006 (7.5)	1.23 (1.14–1.34)	1.25 (1.15–1.35)
Region		
Northeast (ref) (3.9)	1.0	1.0
Midwest (5.1)	0.85 (0.73–0.99)	0.91 (0.79–1.05)
South (9.3)	0.81 (0.70–0.94)	0.89 (0.77–1.03)
West (5.0)	0.80 (0.69–0.92)	0.91 (0.78–1.05)
Physical component score§		
High (ref) (11.7)	1.0	1.0
Low (11.6)	1.17 (1.11–1.24)	1.14 (1.07–1.22)
Mental component score§		
High (ref) (11.6)	1.0	n/a
Low (11.6)	0.88 (0.83–0.94)	
Employment		
Employed (ref) (11.1)	1.0	1.0
Not employed (12.2)	1.33 (1.25–1.41)	1.35 (1.24–1.46)
Marital status		
Married (ref) (13.8)	1.0	1.0
Not married (9.5)	0.78 (0.72–0.83)	0.90 (0.83–0.98)
Time since last checkup		
Never (ref) (0.8)	1.0	1.0
>2 years (2.6)	0.84 (0.64–1.12)	0.78 (0.59–1.03)
≤2 years (19.9)	4.37 (3.34–5.70)	3.59 (2.77–4.65)

Odds ratios determined from weighted sample

†*Adjusted for all variables in table except MCS*

‡*Poor/near-poor (<125% Federal Poverty Level), low (125–<200 % FPL), middle (200–<400% FPL), high (≥400% FPL)*

§*Low = scores less than the median of the study population; high = scores greater than or equal to the median of the study population*

communication with their providers compared to the Other Language-Discordant cohort, thereby leading to higher CRC screening rates.

These findings might be due to differences between the Other Language-Concordant and Other Language-Discordant cohorts. Compared to the Other Language-Concordant group, the Other Language-Discordant respondents were less likely to be Hispanic (42% vs 61%) and more likely to be white non-Hispanic (23% vs 14%). They were also more likely to have attended college (19% vs 12%) and have a high income (27% vs 17%). Income and education are predictors of preventive health-care use;[25] however, these variables were controlled for in our model.

The discrepancy in CRC screening between the Other Language-Concordant and Other Language-Discordant groups may also be related to other unmeasured differences between the two groups. For example, provider cultural competence[7,13,26,27] and better quality of communication between patients and providers[4,22] are associated with higher CRC screening rates. Similarly, rates are higher with greater patient acculturation[28] and health literacy.[9,27,29] While these variables have been shown to be associated with CRC screening rates, we were unable to measure and include them in our study.

There are additional limitations to this study. It is possible that individuals who do not speak English at home speak English well enough to communicate adequately but answered that someone at their provider's office did not speak their language. These individuals would be misclassified as Other Language-Discordant. As a result, our cohorts may not have appropriately captured patient-provider language barriers. Some suggest that LEP is a better measure of language barriers.[4] To address this, alternatively defined cohorts based on comfort speaking English were used. Results in an unadjusted model were similar to those based on language spoken at home. Therefore, regardless of how we defined language concordance, our results suggest that individuals who are 'other language-concordant' with their providers have lower adherence to CRC screening.

In addition, our definition of being adherent to CRC screening guidelines is conservative and may misclassify some as non-adherent. To address this, multivariate models substituting adherence to current CRC guidelines with other CRC screening outcomes were analyzed. Results showed similar findings; individuals in the Other Language-Discordant groups had higher rates for each of the CRC testing outcomes compared to individuals in the Other-Language Concordant group. Furthermore, we could not identify if FOBT or endoscopy was done for diagnostic purposes due to symptoms or in individuals with higher risk, such as those with family history of CRC, which could overestimate CRC screening rates. We did not control for patient co-morbidities, which may influence the appropriateness of screening. As a proxy for co-morbidities, however, we included physical summary health status scores (PCS) in our multivariate model.[30]

Similar to prior studies, our results suggest that speaking a language other than English at home is associated with lower CRC screening. In addition, in our adjusted model we found that individuals who do not speak English at home and do not have anyone at their provider's who speaks their preferred language had CRC screening rates comparable to English speakers, while those who do not speak English at home and

have someone at their provider's who speaks their preferred language, had lower rates. These findings may be related to unmeasured differences between the two cohorts, including patient characteristics, provider cultural competence, patient acculturation, the quality of patient-provider communication, and the level of patient health literacy. Our results suggest that providers should especially promote the importance of CRC screening to non-English speaking patients, but that patient-provider language barriers do not fully account for lower CRC screening in patients who do not speak English at home.

Acknowledgements: *This research was unfunded. Results were presented in poster format at the national meetings of the American College of Preventive Medicine, February 2010, Washington DC and AcademyHealth, June 2010, Boston MA.*

Conflict of Interest: *None disclosed.*

Corresponding Author: *Amy Linsky, MD, MSc; Section of General Internal Medicine, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, GIM, Boston, MA, USA (e-mail: Amy.Linsky@bmc.org).*

REFERENCES

1. US Census Bureau, American FactFinder. Available at: http://factfinder.census.gov/ Accessed August 19, 2010.
2. National Healthcare Disparities Report, 2008. Available at: http://www.ahrq.gov/qual/qrdr08.htm Accessed August 19, 2010.
3. **Fiscella K, Franks P, Doescher MP, Svaer BG.** Disparities in health care by race, ethnicity and language among the insured. Med Care. 2002;40:52–9.
4. **Flores G.** The impact of medical interpreter services on the quality of health care: a systematic review. Med Care Res Rev. 2005;62(3):255–99.
5. **Lavizzo-Mourey R.** Improving quality of US health care hinges on improving language services. JGIM. 2007;22(Suppl 2):279–80.
6. **Collins KS, Hughes DL, Doty MM, Ives BL, Edwards JN, Tenney K**. Diverse communities, common concerns: assessing health care quality for minority americans. The Commonwealth Fund, 2002. Available at: http://www.cmwf.org publication number 523 Accessed August 19, 2010.
7. **Wilson E, Chen AH, Grumbach K, Wang F, Fernandez A.** Effects of limited English proficiency and physician language on health care comprehension. JGIM. 2005;20:800–6.
8. **Pippins JR, Alegria M, Haas JS.** Association between language proficiency and the quality of primary care among a national sample of insured Latinos. Med Care. 2007;45:1020–5.
9. **Chan KS, Keeler E, Schonlau M, Rosen M, Mangione-Smith R.** How do ethnicity and primary language spoken at home affect management practices and outcomes in children and adolescents with asthma? Arch Pediatr Adolesc Med. 2005;159:283–9.
10. **Cheng EM, Chen A, Cunningham W.** Primary language and receipt of recommended health care among Hispanics in the United States. JGIM. 2007;22(Suppl 2):283–8.
11. **Hu DJ, Covell RM.** Health care usage by Hispanic outpatients as a function of primary language. West J Med. 1986;144(4):490–3.
12. **Ponce NA, Hays RD, Cunningham WE.** Linguistic disparities in health care access and health status among older adults. JGIM. 2006;21:786–91.
13. **Harmsen JAM, Bernsen RMD, Bruijnzeels MA, Meeuwesen L.** Patients' evaluation of quality of care in general practice: what are the cultural and linguistic barriers? Pat Educ Couns. 2008;72:155–62.
14. **Whitlock EP, Lin JS, Liles E, Beil T, Fu R.** Screening for colorectal cancer: a targeted, updated systematic review for the US Preventive Services Task Force. Ann Intern Med. 2008;149:638–58.
15. **Jamal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ.** Cancer statistics, 2007. CA Cancer J Clin. 2007;57(1):43–66.
16. **Peterson NB, Murff HJ, Ness RM, Dittus RS.** Colorectal cancer screening among men and women in the United States. J Womens Health. 2007;16(1):57–65.
17. **Natale-Pereira A, Marks J, Vega M, Mouzon D, Hudson S, Salas-Lopex D.** Barriers and facilitators for colorectal cancer screening practices in the Latino community: perspectives from community leaders. Cancer Control. 2008;15(2):157–165.
18. **Diaz JA, Roberts MB, Goldman RE, Weitzen S, Eaton CB.** Effect of language on colorectal cancer screening among Latinos and non-Latinos. Cancer Epidemiol Biomarkers Prev. 2008;17(8):2169–73.
19. **Nguyen GT, Barg FK, Armstrong K, Holmes JH, Hornik RC.** Cancer and communication in the health care setting: experiences of older Vietnamese immigrants, a qualitative study. JGIM. 2007;23(10):45–50.
20. **Wee CC, McCarthy EP, Phillips RS.** Factors associated with colon cancer screening: the role of patient factors and physician counseling. Prev Med. 2005;41:23–9.
21. **Guerra CE, Schwartz JS, Armstrong K, Brown JS, Halbert CH, Shea JA.** Barriers of and facilitators to physician recommendation of colorectal cancer screening. JGIM. 2007;22(12):1681–8.
22. **Dunn AS, Shridharani KV, Lou W, Bernstein J, Horowitz CR.** Physician-patient discussions of controversial cancer screening tests. Am J Prev Med. 2001;20:130–4.
23. **Carcaise-Edinboro P, Bradley CJ.** Influence of patient-provider communication on colorectal cancer screening. Med Care. 2008;46:738–45.
24. **Ware JE, Kosinski M, Keller SD.** A 12-Item Short-Form health survey. Med Care. 1996;34(3):220–33.
25. **Hershey JC, Luft HS, Gianaris JM.** Making sense out of utilization data. Med Care. 1975;13(10):838–854.
26. **Wolf MS, Baker DW, Makoul G.** Physician-patient communication about colorectal cancer screening. JGIM. 2007;22(11):1493–9.
27. **Safran DG, Taira DA, Rogers WH, Kosinski M, Ware JE, Tarlov AR.** Linking primary care performance to outcomes of care. J Fam Pract. 1998;47:213–220.
28. **Etzioni DA, Ponce NA, Babey SH, et al.** A population-based study of colorectal cancer test use: results from the 2001 California Health Interview Survey. Cancer. 2004;101:2523–32.
29. **Sudore RL, Landefeld CS, Perez-Stable EJ, Bibbins-Domingo K, Williams BA, Schillinger D.** Unraveling the relationship between literacy, language proficiency, and patient-physician communication. Patient Educ Couns. 2009;75:398–402.
30. **Kazis LE, Nethercot VA, Ren XS, Lee A, Selim A, Miller DR.** Medication effectiveness studies in the United States Veterans Administration health care system: a model for large integrated delivery systems. Drug Dev Res. 2006;67:217–226.

J Thromb Thrombolysis (2011) 32:242–266
DOI 10.1007/s11239-011-0592-7

Highlights from the III International Symposium of Thrombosis and Anticoagulation (ISTA), October 14–16, 2010, São Paulo, Brazil

Renato D. Lopes · Richard C. Becker · John H. Alexander · Paul W. Armstrong · Robert M. Califf · Mark Y. Chan · Mark Crowther · Christopher B. Granger · Robert A. Harrington · Elaine M. Hylek · Stefan K. James · E. Marc Jolicoeur · Kenneth W. Mahaffey · L. Kristin Newby · Eric D. Peterson · Karen S. Pieper · Frans Van de Werf · Lars Wallentin · Harvey D. White · Antonio C. Carvalho · Roberto R. Giraldez · Helio P. Guimaraes · Helena B. Nader · Renato A. K. Kalil · Joyce M. A. Bizzachi · Antonio C. Lopes · David A. Garcia

Published online: 6 May 2011

Abstract To discuss and share knowledge around advances in the care of patients with thrombotic disorders, the Third International Symposium of Thrombosis and Anticoagulation was held in São Paulo, Brazil, from October 14–16, 2010. This scientific program was developed by clinicians for clinicians, and was promoted by four major clinical research institutes: the Brazilian Clinical Research Institute, the Duke Clinical Research Institute of the Duke University School of Medicine, the Canadian VIGOUR Centre, and the Uppsala Clinical Research Center. Comprising 3 days of academic presentations and open discussion, the symposium had as its primary goal to educate, motivate, and inspire internists, cardiologists,

The Third International Symposium of Thrombosis and Anticoagulation (ISTA) was sponsored by unrestricted grants from Alexium, AstraZeneca, Boehringer Ingelheim, Bristol-Myers Squibb, CSL Behring, Daiichi Sankyo, Leo Pharma, and Sanofi-aventis. The symposium was organized by Meeting Eventos, and the chairmen of the meeting thank Cecilia Jahnel, Fernanda Prestes, and Daniela Rodrigues for all of their assistance in organizing the event.

R. D. Lopes (✉) · R. C. Becker · J. H. Alexander · R. M. Califf · C. B. Granger · R. A. Harrington · K. W. Mahaffey · L. K. Newby · E. D. Peterson · K. S. Pieper
Duke Clinical Research Institute, Duke University Medical Center, Box 3850, 2400 Pratt Street, Room 0311, Terrace Level, Durham, NC 27705, USA
e-mail: renato.lopes@duke.edu

R. D. Lopes · A. C. Carvalho · H. P. Guimaraes · H. B. Nader · A. C. Lopes
Paulista School of Medicine, Federal University of São Paulo, São Paulo, Brazil

P. W. Armstrong
University of Alberta, Edmonton, AB, Canada

M. Y. Chan
National University Heart Center, Singapore, Singapore

M. Crowther
McMaster University, Hamilton, ON, Canada

E. M. Hylek
Boston University, Boston, MA, USA

S. K. James · L. Wallentin
Uppsala University, Uppsala, Sweden

E. M. Jolicoeur
Montreal Heart Institute, Université de Montréal, Montreal, QC, Canada

F. Van de Werf
University Hospitals Leuven, Leuven, Belgium

H. D. White
Green Lane Cardiovascular Research Unit, Auckland, New Zealand

R. R. Giraldez
São Paulo University Medical School, São Paulo, Brazil

R. A. K. Kalil
Federal University of Health Science of Porto Alegre, Porto Alegre, Brazil

J. M. A. Bizzachi
State University of Campinas, São Paulo, Brazil

D. A. Garcia
University of New Mexico Health Sciences Center, Albuquerque, NM, USA

hematologists, and other physicians by convening national and international visionaries, thought-leaders, and dedicated clinician-scientists. This paper summarizes the symposium proceedings.

Keywords Thrombosis · Antithrombotic therapy · Guidelines · Clinical research

Introduction

Importance of thrombosis

Venous and arterial thrombosis remains the most frequent cause of death in western countries. Cardiovascular disease, including heart attack and stroke, accounts for more than 50% of deaths (http://www.cdc.gov/nchs/fastats/deaths.htm). Additionally, the presence of thromboembolism is an adverse prognostic indicator in patients with cancer, which is the second most common cause of death. As a result, there is great interest in the development of novel anticoagulant agents designed to reduce the risk of first or recurrent thrombotic event while minimizing the risk of bleeding. Arterial thrombosis is generally due to platelet activation occurring at sites of vascular injury in high-flow and high-sheer vessels. Generally, antiplatelet agents are preferred for primary or secondary prevention of arterial thrombosis because they inhibit platelet activation induced by platelet binding at sites of vascular injury and mediated by von Willebrand factor. Recent interest has focused on the development of new and more potent antiplatelet agents with special characteristics including rapid on- and off-set of action, shorter half-lives, and more potent inhibition of specific self-surface receptors including the thrombin receptor.

Venous thrombosis is generally thought to be due to activation of soluble coagulation proteins in low-flow areas of the venous system. There are some parallels in the left atrium of patients with atrial fibrillation (AF), suggesting that treatments that are effective for prevention of venous thrombosis will also be effective for prevention of systemic embolization in patients with AF. Traditional agents for prevention and treatment of venous thrombosis include heparins, low-molecular-weight heparins (LMWH), pentasaccharides, and a variety of parenteral anticoagulants used infrequently in specific circumstances such as patients with heparin-induced thrombocytopenia. Long-term therapy has traditionally been provided by warfarin administered to achieve an international normalized ratio (INR) of 2.0–3.0. The limitations of warfarin—including drug and food interactions, variability within and between patients in dosing requirements, a narrow therapeutic window, and the need for frequent INR monitoring—have led to the development of novel agents that lack some or all of these characteristics. Dabigatran and rivaroxaban are two agents that have been approved for several indications. Dabigatran recently was approved in Canada and the United States for prevention of systemic embolization in patients with AF. These agents, if proven safe in phase IV studies, offer significant advantages over warfarin for prevention of systemic embolization. They are also the subject of studies for secondary prevention of venous thrombosis. In this setting, efficacy of both agents is comparable to warfarin.

Intensification of antithrombotic therapy has a cost. There is clear evidence that bleeding rates increase as patients are treated with more aggressive antithrombotic regimens. Thus, when compared with warfarin alone, bleeding risks increase in patients treated with aspirin and warfarin, and further increase in patients treated with so-called "triple therapy." Risks of bleeding will undoubtedly be even higher in patients who are treated with "quadruple therapy," as novel antiplatelet and antithrombotic agents are brought to market.

There is also evidence that a therapeutic effect can be achieved at lower doses of antithrombotic medications than are currently employed for many indications. Thus, prophylactic doses of pentasaccharide are as effective as therapeutic doses of enoxaparin for prevention of thrombotic and other vascular complications in patients with unstable coronary syndromes. At prophylactic doses, fondaparinux produces less bleeding than enoxaparin, suggesting it may be a preferred agent for treatment in this setting. The pentasaccharide study highlights current thoughts suggesting that "de-intensification" should be considered in selected patients because currently available antithrombotics may maintain their "therapeutic effect" at levels that are associated with a lower rate of "toxicity," predominantly bleeding.

In summary, cardiovascular disease remains a leading cause of death. Significant resources have been invested in the design and evaluation of novel antithrombotic agents, which are now being evaluated for prevention of both first and recurrent thrombotic events in high-risk patients. Demonstration that intensification of anticoagulation is associated with enhanced bleeding risk has led to studies that attempt to de-intensify antithrombotic therapy. Novel agents offer the hope of simplicity of treatment with reduced toxicity; however, their safety must be proven in large patient groups.

ISTA

To discuss and share knowledge around advances in the care of patients with thrombotic disorders, the Third International Symposium of Thrombosis and Anticoagulation (ISTA) was held in São Paulo, Brazil, from October

14–16, 2010. This scientific program was developed by clinicians for clinicians, and was promoted by four major clinical research institutes: the Brazilian Clinical Research Institute (BCRI), the Duke Clinical Research Institute (DCRI) of the Duke University School of Medicine, the Canadian VIGOUR Centre (CVC), and the Uppsala Clinical Research Center (UCR). It was also supported by the Brazilian Societies of Internal Medicine, Cardiology, Intervention Cardiology, Heart Failure, Nephrology, Intensive Care Medicine, Hematology, Oncology, and Vascular Surgery, by the Latin American Group of Thrombosis and Hemostasis, and by the Anticoagulation Forum from the United States. The chairmen of the meeting were Dr. Renato D. Lopes and Dr. Richard C. Becker, both from Duke University School of Medicine and the Duke Clinical Research Institute, and Dr. David Garcia from the University of New Mexico.

Comprising 3 days of academic presentations and open discussion, the symposium had as its primary goal to educate, motivate, and inspire internists, cardiologists, hematologists, and health care providers by convening national and international visionaries, thought-leaders, and dedicated clinician-scientists to review the scientific evidence in the area of thrombosis. The following is a summary of the symposium proceedings.

Platelet biology

Platelet biology and an advanced understanding of fundamental concepts governing the behavior of platelets, both in terms of pathologic thrombotic events and the support of normal hemostasis, comprise a vital part of identifying targets for drug development and achieving optimal patient care. There are four constructs or functional themes of importance: platelet aggregation, platelet support of coagulation, platelet support of vascular integrity, and platelet support of vascular repair.

The initiation of coagulation is characterized by the assembly of coagulation proteins on tissue factor-bearing cells. This is followed by thrombin generation and, if of sufficient quantity to cause platelet activation, platelet aggregation, assembly of coagulation proteins, and a "burst" of thrombin generation with subsequent clot propagation.

Platelet activation and aggregation occurring at a site of vessel wall injury is characterized by three distinct populations of platelets. The first population is characterized by expression of ligand receptors, which in turn facilitate platelet aggregation. The second is characterized by the expression of phosphatidylserine with support coagulation protein assembly and thrombin generation. The third consists of a population of platelets with predominantly paracrine effects that are required for the important stage of vessel wall healing. This latter population of platelets has been underappreciated in considering the potential effects of long-term, robust platelet inhibition with pharmacological therapy.

Several recent observations shed new light on the important interface between platelets and coagulation protein activation within the developing thrombus. Specifically, the release of platelet polyphosphates has been shown to activate factors XI and XII, facilitating thrombin generation. More recent information also highlights the role of polyphosphates, factor XI, and factor XII as triggers of thrombosis that are not required for normal hemostasis. These observations will likely prompt increasing interest in new targets with the theoretical potential to uncouple thrombosis and hemostasis.

The importance of platelets in both a reparative capacity and as facilitators of inflammation highlights their pleotropic capabilities. Despite being anuclear cells, megakaryocytes within the bone marrow respond to a variety of signals, potentially being reprogrammed in the presence of specific conditions. In addition, the recognition that activated human platelets splice pre-mRNA into mature transcripts supports a highly dynamic capability. Whether platelet antagonists can influence either programming at the level of the megakaryocyte or peripheral circulation splicing of pre-mRNA will require further investigation. It is becoming increasingly clear that platelets no longer can be viewed as passive bystanders to vascular events and systemic conditions.

Measures of platelet function

For the last several decades, measurement of platelet function has been used primarily for diagnosis of intrinsic deficiencies of platelet hemostatic capacity. However, more recent work has focused on platelet function testing as a pharmacodynamic measure of response to platelet-directed therapy. In the treatment of atherothrombosis, inhibition of platelet activation and aggregation plays a central role in attenuating thrombus formation and propagation. Such antagonism of the atherothrombotic process is vital for secondary prevention in patients with acute coronary syndromes (ACS) and after percutaneous coronary intervention (PCI).

Two antiplatelet medications used commonly in these populations are aspirin and clopidogrel. Platelet function testing has documented substantial variability in the pharmacodynamic response to both medications; however, the prevalence and clinical impact of this variability remain largely unknown.

Several methods for assessing platelet responsiveness to clopidogrel or aspirin are available. Light-transmission aggregometry (LTA) is the historical "gold standard" for evaluation of the pharmacodynamic response to platelet-directed medications. The major disadvantages are: (1) increased processing time per sample because of the need to generate platelet rich plasma, and (2) increased inter-operator variability. The vasodilator-activated phosphoprotein (VASP) test measures the intracellular platelet response to medications inhibiting the platelet $P2Y_{12}$ receptor. Like LTA, it is a time-consuming, laboratory-based test that is technically demanding.

Newer point-of-care whole blood aggregation tests are now available. Of these, the VerifyNow® (Accumetric, Inc., San, Diego, CA) and Multiplate® (Verum Diagnostica, Munich, Germany) tests have undergone extensive clinical validation. In the Do Platelet Function Assays Predict Clinical Outcomes in Clopidogrel Pretreated Patients Undergoing Elective PCI (POPULAR) study comparing the predictive value of different platelet function tests for thrombotic and bleeding outcomes following elective PCI, the VerifyNow test demonstrated a c-statistic comparable to LTA in its ability to discriminate future thrombotic outcomes. Similarly, the Multiplate test performed well in a large German multicenter study investigating the relationship between its adenosine diphosphate (ADP) test and thrombotic outcomes after PCI.

However, the ability of existing platelet function tests to predict bleeding outcomes is more limited. To date, the association between platelet function measurements and future bleeding outcomes has been equivocal—this remains a key limitation of platelet function testing. Another key limitation of currently available platelet function tests is their inability to reliably report on the composite effect of multiple antiplatelet agents acting via different pathways. A further unresolved question is the ability of a platelet function testing-guided strategy to improve clinical outcomes. Although the recently completed Gauging Responsiveness with a VerifyNow Assay—Impact on Thrombosis and Safety (GRAVITAS) trial did not demonstrate an improvement in clinical outcomes with double-dose clopidogrel in patients with clopidogrel hypo-responsiveness identified using the VerifyNow system, other ongoing trials employing more potent $P2Y_{12}$ inhibitors will provide greater clarity on the clinical utility of platelet function testing.

Proton-pump inhibitor (PPI)–clopidogrel interactions: reality or myth?

In patients with ACS, current clinical practice guidelines recommend the use of dual antiplatelet therapy with aspirin and $P2Y_{12}$ inhibition. Aspirin, particularly at higher doses, leads not only to platelet inhibition by the effect on thromboxane A2 but also to effects on the gastric mucosa through the inhibition of prostacyclin. This results in an increased risk of peptic ulcer and gastric bleeding. Addition of clopidogrel to aspirin further increases the risk of adverse gastric bleeding events. To reduce this risk, treatment with proton-pump inhibitors (PPI) is routinely used in patients with previous peptic ulcer and often in patients with risk factors for gastric bleeding such as acute care. In fact, both clopidogrel and PPIs are among the most frequently prescribed pharmacological agents worldwide.

The most important mechanism for a poor response to clopidogrel is variable generation of the active metabolite. Approximately 85% of a clopidogrel dose is hydrolysed by esterases to an inactive metabolite. The remaining clopidogrel is available to be converted to the active metabolite in a process requiring two sequential cytochrome P450 (CYP)-dependent steps with CYP2C19 in both steps. A genetically determined reduced function allele of CYP2C19 slows clopidogrel metabolism, which leads to lower levels of the clopidogrel active metabolite and a lower pharmacodynamic platelet inhibitory effect.

Because some PPIs are known to be strong inhibitors of CYP2C19 activity, it is reasonable to believe that PPI may reduce the clinical response to clopidogrel. Controversy remains over whether this treatment interaction is clinically meaningful.

In November 2009, the U.S. Food and Drug Administration (FDA) issued a warning that concomitant use of omeprazole and clopidogrel should be avoided and that other drugs that reduce stomach acid do not interfere with the anti-clotting activity of clopidogrel. The European Medicines Agency (EMEA) extended the warning to discourage concomitant use of all PPIs unless absolutely necessary. These recommendations were based on pharmacokinetic/pharmacodynamic and observational studies. Well-performed studies have shown that the mean plasma concentration of the clopidogrel active metabolite is lower in patients treated with omeprazole in combination with clopidogrel than in patients treated with clopidogrel alone, also with a 600-mg loading and 150-mg maintenance dose. Pharmacodynamic studies have confirmed the reduction of platelet reactivity.

Whether treatment with PPIs affects cardiovascular outcome in patients receiving clopidogrel has been unclear. Several small observational studies showed a significant association between PPI use and cardiovascular risk, whereas propensity-matched studies and substudies of large randomized trials such as the Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition with Prasugrel-Thrombolysis in Myocardial Infarction (TRITON-TIMI) 38 and the Platelet

Inhibition and Patient Outcomes (PLATO) study revealed no association. A recently performed meta-analysis including 159,138 patients from 25 studies found an association of PPIs with reduction in gastric bleeding events and a higher risk of stent thrombosis but no association with the risk of death. One randomized trial, the Clopidogrel and the Optimization of Gastrointestinal Events Trial (COGENT), studied the effect of omeprazole versus placebo in patients treated with dual antiplatelet therapy. Although the trial was stopped prematurely for financial reasons, 3,637 patients were enrolled. The trial showed a 66% relative reduction in gastrointestinal events but no effect of omeprazole on cardiovascular events.

In summary, the totality of data suggests that a pharmacokinetic and pharmacogenetic clopidogrel–PPI interaction via CYP2C19 is real but that concomitant use of PPIs has minimal or no clinical consequence in low–medium-risk patients on long-term treatment. A small but clinically meaningful interaction with PPI in ACS patients at high ischemic risk in the acute settings cannot be excluded. Treatment with other potent $P2Y_{12}$ receptor inhibitors, such as prasugrel and ticagrelor, is not associated with an interaction with PPIs and could be considered in patients at high risk for ischemic events. In patients at risk for peptic ulcer, treatment with effective gastric protection, including PPIs, should not be withheld.

New antiplatelet agents under development

Current management of ACS includes risk stratification by clinical findings and the use of electrocardiographic and biochemical markers. It is recommended that all patients with an established diagnosis of ACS receive immediate antithrombotic treatment with dual platelet inhibition (aspirin and a $P2Y_{12}$ inhibitor) plus intravenous or subcutaneous anticoagulation. In addition, patients should also receive beta-blockers, statins, and, frequently, angiotensin-converting enzyme (ACE) inhibitors. The majority of patients hospitalized for ACS are rapidly admitted to a catheterization laboratory for identification of the culprit lesion, followed by balloon dilatation and stenting if feasible. At discharge, it is generally recommended that patients receive long-term secondary prevention with a combination of aspirin, beta-blockers, statins, ACE inhibitors, and $P2Y_{12}$ inhibitors for at least 1 year. However, despite these measures, there is still a 10% risk of death, reinfarction, or stroke during the year following discharge. The magnitude of this risk varies among patient populations, with the highest risk in older patients and those with diabetes mellitus, previous myocardial infarction (MI), cardiac or renal dysfunction, manifestations of atherosclerotic disease, or multi-vessel coronary artery disease (CAD). If current therapeutic approaches for ACS are to be improved, greater focus will be needed on these high-risk groups.

New therapies currently under development aim to prevent further progression of thrombosis and atherosclerosis and to correct underlying metabolic disturbances (e.g., diabetes and dyslipidemia). The primary challenge in preventing and managing ACS, both now and in the future, will be to tailor treatments for each patient, taking into consideration patient characteristics, comorbidities, underlying short- and long-term risk factors, and expected individual responses to different medications. These ambitions will likely place a substantial burden on global health care resources and may ultimately require prioritization among several treatment alternatives.

Platelet inhibition has been a mainstay in the prevention of MI and death in patients with ACS for approximately 20 years. Aspirin therapy yields consistent inhibition of platelet thromboxane A_2 release. However, inhibiting this pathway only modestly attenuates platelet activation without any influence on ADP-induced platelet activation. Aspirin treatment reduces the relative risk of MI and death by 30–50% compared with placebo in patients with ACS. However, aspirin alone has no convincing effect on prevention of stent thrombosis. Therefore, other pathways need to be inhibited in the highly prothrombotic environment of ACS. The $P2Y_{12}$ receptor plays a major role in the ADP-mediated amplification of platelet response regardless of the stimulus. Clopidogrel and prasugrel are thienopyridine pro-drugs acting on this receptor by almost identical active metabolites that irreversibly bind to the receptor. Slow and variable active metabolite generation leads to clopidogrel having a slow onset of action and wide inter-individual variability in pharmacodynamic response. However, as shown in the Clopidogrel in Unstable Angina to Prevent Recurrent Events (CURE) trial, compared with aspirin alone, clopidogrel provided a relative 20% reduction in death, MI, or stroke at a median of 9 months of treatment. Despite variability in platelet responsiveness, the Clopidogrel Optimal Loading Dose Usage to Reduce Recurrent Events/Optimal Antiplatelet Strategy for Interventions (CURRENT/OASIS) 7 trial—a 2×2 factorial randomized comparison of standard-dose (300 mg load/75 mg daily) versus higher-dose (600 mg load/150 mg daily for 6 days, then 75 mg daily) clopidogrel and lower-dose (75–81 mg) versus higher-dose (325 mg) aspirin treatment—failed to show superiority for the higher-dose clopidogrel regimen. However, in a subgroup analysis of PCI-treated patients, there was a substantial reduction in stent thrombosis in patients treated with the higher-dose clopidogrel regimen.

Generation of the active metabolite of prasugrel is more efficient than for clopidogrel, resulting in more rapid onset

of action, more pronounced platelet inhibition, and no clinically important variability in response. In the setting of these pharmacokinetic and pharmacodynamic features, prasugrel treatment resulted in a 20% reduction in death, MI, or stroke and a halving of the risk of stent thrombosis compared with clopidogrel in the TRITON-TIMI 38 trial of patients undergoing a planned PCI procedure. However, both the addition of clopidogrel to aspirin and the use of prasugrel instead of clopidogrel were associated with significant increases in major bleeding in CURE and TRITON-TIMI 38, respectively.

Ticagrelor, the first reversibly binding oral $P2Y_{12}$ receptor antagonist, does not require metabolic activation, has a rapid onset of action, and can disassociate from the receptor, permitting restoration of platelet function without the need for production of new platelets. In pharmacodynamic studies, ticagrelor demonstrated greater, more rapid, and more consistent ADP-induced platelet inhibition compared with clopidogrel and more rapid offset of action following cessation of therapy. In the PLATO study, 18,624 patients with ACS were randomized within 24 h after symptom onset to ticagrelor versus clopidogrel. The results showed a 16% relative reduction of the composite of cardiovascular death, MI, or stroke, a 22% reduction in total mortality, and a 33% reduction in definite stent thrombosis. Ticagrelor was not associated with an increase in overall bleeding, but, during long-term treatment, there was more non-procedural bleeding with ticagrelor.

Currently, elinogrel, a reversibly binding competitive $P2Y_{12}$ receptor antagonist for both intravenous and oral administration, is under evaluation. In a recently presented phase II trial (Novel Intravenous and Oral P2Y12 Inhibitor in Non-Urgent PCI [INNOVATE-PCI]), elinogrel was associated with a slight dose-related increase in total bleeding without a clear signal for reduction in ischemic events compared with clopidogrel.

Other targets for platelet inhibition are also under investigation (e.g., the protease-activated receptor 1 [PAR-1]). Preclinical and phase II studies suggest that consistent and high levels of PAR-1 inhibition may have a beneficial antithrombotic effect with minimal increase in bleeding. Phase III studies of the selective PAR-1 inhibitor, vorapaxar, are currently underway, both in ACS and chronic CAD.

In conclusion, several new alternatives providing more rapid and consistent platelet inhibition than clopidogrel are currently being explored for routine treatment of patients with ACS. These new treatments seem to provide additional benefits to the patients without unacceptable increases in the risk of bleeding if used appropriately. Within the next few years, even more treatment alternatives might be available to further improve outcomes of the large patient population with ACS.

Novel parenteral anticoagulants

Despite its limitations, unfractionated heparin (UFH) remains a commonly used parenteral anticoagulant in clinical practice. The major limitations of UFH include an unpredictable pharmacodynamic response, associated off-target effects, and the need for pharmacodynamic monitoring. Lack of pharmacologic specificity is another limitation. As such, the scientific community has moved toward using novel anticoagulants that target singular proteases within the coagulation system.

Bivalirudin, a direct thrombin inhibitor with a short circulating half-life, has recently shown good clinical efficacy with less bleeding compared with either UFH or LMWH. The synthetic pentasaccharide, fondaparinux, is a specific, indirect inhibitor of factor Xa. Despite, the convenience of a once-a-day subcutaneous injection for the management of patients with ACS, the need for supplemental UFH during transition to the catheterization laboratory limits its wider adoption in clinical practice. Concerns also remain over the propensity for equipment-associated thrombosis, as well as the absence of a reliable antidote to reverse its anticoagulant effect. Another factor Xa inhibitor, otamixaban, has demonstrated early safety as a parenteral anticoagulant in the catheterization laboratory; in the phase II Otamixaban in Comparison to Heparin in Subjects Undergoing Non-Urgent Percutaneous Coronary Intervention (SEPIA-PCI) trial, equipment-associated thrombosis occurred at a similar rate in both otamixaban and UFH-treated patients.

In response to existing limitations of approved parenteral anticoagulants, REG1 was designed to achieve rapid inhibition of factor IXa with active, antidote-mediated reversibility. This drug-antidote construct is now undergoing late phase II testing in patients with ACS.

Vitamin K antagonists

The vitamin K antagonists (VKAs) have been the only anticoagulants available for oral use since their first administration to a patient more than 50 years ago. The mechanism of action of this drug class is complex; they achieve their anticoagulant effect by inducing the synthesis of dysfunctional forms of factors II, VII, IX, and X. The target of VKAs is the enzyme vitamin K epoxide reductase (VKOR). Single nucleotide polymorphisms in the gene that codes for this enzyme (as well as single-nucleotide polymorphisms [SNPs] in the genes that encode CYP2C9) can render the patient more (or, in some cases, less) sensitive to warfarin; thus, common genetic variations, along with factors such as sex, age, and weight, lead to significant (and sometimes unpredictable) inter-individual variability in the

dose required to achieve the targeted anticoagulant effect. Even patients whose dose has been determined through titration and adjustment can experience clinically relevant sudden changes in their anticoagulant effect because of interactions with diet (mostly due to variation in vitamin K intake) or other medications (especially drugs that interact with the cytochrome P450 system). These features, along with the narrow therapeutic index, slow onset, and long pharmacodynamic half-life characteristic of VKAs, have created challenges for clinicians and patients alike.

Despite the undesirable attributes of VKAs, they have proven to be extremely effective in the prevention of AF-related stroke, recurrent venous thromboembolism (VTE), and other unwanted clinical events. Recently, the inconvenience of VKAs has been reduced by the opportunity for patient self-testing, but self-testing does not necessarily make VKAs safer or more effective than they are in the context of a dedicated system of anticoagulation management. Indeed, the safety of VKAs has improved with the advent of dedicated anticoagulation management services and the application of evidence-based strategies to reverse the VKA anticoagulant effect in bleeding patients. Going forward, it is likely that VKA use will decrease, but not disappear, once new oral anticoagulant agents become available.

ACS with ST-segment elevation: guidelines perspective on antithrombotic therapy

Many anticoagulants and antiplatelet agents are now available for the treatment of ACS patients. In patients with ST-segment elevation MI (STEMI) undergoing primary PCI, clopidogrel and UFH or bivalirudin (a direct antithrombin agent) are the most frequently used agents. For patients treated with lytic therapy, UFH, enoxaparin, and fondaparinux (with streptokinase only) are used as anticoagulant co-therapy. Clopidogrel is also given routinely with lytic agents to patients under age 75 years. No reliable data are available in patients aged >75 years.

Clopidogrel is widely used as an adjunctive therapy for primary PCI and has also been shown to be beneficial in patients treated with fibrinolytic therapy (Clopidogrel as Adjunctive Reperfusion Therapy [CLARITY] study; Clopidogrel and Metoprolol in Myocardial Infarction Trial [COMMIT/CCS-2]). However, its limitations—particularly slow onset of action, variability in response, and irreversible binding to the P2Y12 receptor—create challenges for STEMI care. Prasugrel, approved for use in Europe in 2009 and in the U.S. in 2009, is a third-generation thienopyridine and has a similar mechanism of action to clopidogrel but superior pharmacokinetic characteristics. The greater efficacy of prasugrel over clopidogrel in the TRITON-TIMI 38 trial was particularly evident in patients with STEMI, all of whom underwent primary PCI. However, prasugrel was associated with an increased risk of major bleeding, although in the STEMI population, there was no increase in life-threatening bleeding compared with clopidogrel.

Several novel antiplatelet therapies are currently in clinical development or have only recently been approved. The PLATO study demonstrated that ticagrelor reduced the incidence of death, MI, or stroke by 16% and of cardiovascular death by 22% compared with clopidogrel in STEMI patients. With ticagrelor, there was an increased risk of non-coronary artery bypass graft (CABG) bleeding complications. Also, cangrelor is an intravenous, fast, direct-acting, and reversible $P2Y_{12}$ inhibitor. No significant differences with clopidogrel could be demonstrated in ACS patients in the Cangrelor versus Standard Therapy to Achieve Optimal Management of Platelet Inhibition (CHAMPION) PCI trial.

It is likely that prasugrel and ticagrelor will be recommended in the guidelines for STEMI patients undergoing primary PCI. Because these agents have not been tested prospectively with lytic agents, clopidogrel will remain the recommended ADP antagonist with lytic therapy.

Obviously, there is no role for new oral anticoagulants such as rivaroxaban and apixaban in the acute reperfusion phase of STEMI. Whether these agents may prevent recurrent ischemic events afterwards is unknown.

Antithrombotic therapy in ACS with non-ST-segment elevation

There are three antithrombotic agents to choose from: enoxaparin, fondaparinux, and bivalirudin. All have been shown to be superior to UFH. However, there are a number of considerations in choosing an antithrombotic agent for non-ST-segment elevation ACS. These include ischemic risk, bleeding risk, whether an invasive or conservative strategy will be employed, time to catheterization (<12 h vs. >12 h), whether drugs will be switched, whether the patient is aged <75 or ≥75 years, and the patient's renal function.

Major bleeding is strongly associated with subsequent mortality and ischemic events and, many believe, is at least as important as reinfarction. Most bleeding complications are iatrogenic, attributable to femoral artery access for PCI, and related to the use of potent antiplatelet and antithrombin medications. The incidence of bleeding is affected by the choice of anticoagulant and overdosing.

Enoxaparin has been shown in trials of over 22,000 patients to reduce death and MI by 20% and to have similar outcomes as compared with UFH when a conservative strategy is employed, but its use is associated with a modest increase in bleeding when an invasive strategy is employed.

There is also a large clinical experience for bivalirudin in non-ST-segment elevation ACS. Over 20,000 patients have been randomized in trials showing that major bleeding is reduced by about 50% with no increase in ischemia compared with UFH.

Crossover of antithrombotics

The patients in the Superior Yield of the New Strategy of Enoxaparin, Revascularization, and Glycoprotein IIb/IIIa Inhibitors (SYNERGY) trial who crossed over between UFH and enoxaparin had an increase in bleeding complications. Crossover occurred at various times through the study period, at times in response to clinical or clinician perception. In a secondary analysis from this study, results indicated a significant association between crossover from enoxaparin to UFH and TIMI bleeding but not in the other direction, and no crossover association was found in death or MI.

Switching from UFH or enoxaparin to bivalirudin in the Acute Catheterization and Urgent Intervention Triage Strategy (ACUITY) trial was not associated with an increased risk for ischemic events. Furthermore, switching to bivalirudin provided patients with a 50% reduction in bleeding.

Fondaparinux is an indirect factor Xa inhibitor tested against enoxaparin in the Organization to Assess Strategies in Acute Ischemic Syndromes (OASIS)-5 trial. The ischemia rate was similar to UFH, but severe bleeding complications were significantly reduced with fondaparinux, and long-term mortality and stroke rates were also reduced. Because of a higher rate of catheter thrombosis when fondaparinux alone is used, UFH (85 μ/kg) should be added for patients undergoing PCI.

If there is a very high risk of ischemia, bivalirudin is recommended, or UFH with a IIb/IIIa antagonist added if there is angiographic thrombosis or poor TIMI flow. If there is a low-to-high risk of ischemia, all four agents (fondaparinux, LMWH, UFH, and bivalirudin) are good choices. Bivalirudin is an attractive option if there is an increased risk of bleeding and an early invasive strategy is planned. Fondaparinux is a good option if a conservative strategy is planned.

A number of different anticoagulant strategies can be appropriately selected based on individual risk stratification for ischemia and bleeding.

Biomarkers of thrombosis: where do we stand in 2010?

From a clinical perspective, biomarkers serve three main purposes: to diagnose or exclude a disease diagnosis; to provide information about prognosis or to risk stratify; and, most elusively, to guide treatment decisions. Pulmonary embolism (PE) and acute MI are two acute thrombotic disease entities, often presenting with similar symptoms, for which clinically useful biomarkers of thrombosis have evolved across each of these three domains.

Pulmonary embolism is first classified as high or intermediate/low risk based on hemodynamic and respiratory status. For those who are stable (i.e., intermediate/low risk), the biomarker, D-dimer, is used to exclude the diagnosis of PE and to guide further imaging and/or treatment. Given its exquisite sensitivity, despite low specificity, its negative predictive value is very high, such that further work-up with imaging or treatment is not necessary if the D-dimer concentration is low.

Troponin remains the gold standard for establishing a diagnosis of MI in the setting of clinical symptoms of ischemia. Troponin assays are more sensitive and more specific for myocardial injury than creatine kinase (CK)-MB. However, the increasing clinical availability of high-sensitivity troponin assays that can detect circulating troponin at levels well below the 99th percentile of a normal reference population and can also achieve 10% coefficient of variation (CV) at the 99th percentile is challenging the diagnostic utility of troponin testing for diagnosis of MI. However, the increased sensitivity of these assays is offset by reduced clinical specificity, resulting in low positive predictive value. For example, up to 70% of patients with heart failure, which often co-exists with coronary disease, may present with elevated troponin by high-sensitivity assays. The parameters for diagnostic use of these assays are still being discussed. However, these assays may be particularly useful in early diagnosis/triage in the emergency room, where elevations above the 99th percentile in MI patients are detectable much earlier than with standard assays. Given these challenges, heightened awareness of the relationship of pre-test probability with the occurrence of false-positive (and false-negative) diagnoses will be needed.

In the meantime, systematic efforts to increase the accuracy of physicians' clinical assessments of risk in patients with suspected ACS must be undertaken. In a study from the Canadian ACS 2 Registry, despite the availability of the 12-lead electrocardiogram (ECG) and results of assays for markers of myocardial necrosis, there was little relationship between physician-estimated risk category and that determined from available risk scores, with wide variability in these risk scores within the physician-estimated category. A better alignment between physician-estimated risk and systematically determined risk is critical as this study also showed that physicians, overall, treat patients whom they judge as being at higher risk more aggressively with both coronary procedures and medications.

Even with a systematic approach to risk stratification, novel or existing biomarkers may be useful in refining prognosis or guiding treatment selection. Many biomarkers of thrombosis and inflammation have been identified, but rarely have studies considered more than a few biomarkers simultaneously, and few have made the translation from biomarker of risk to biomarker for stratified application of treatments. Troponin testing is the cardiovascular biomarker that best exemplifies this feature. In studies mostly done with older assays, troponin identified high-risk populations that were most likely to benefit from glycoprotein (GP) IIb/IIIa inhibitors, LMWH, and a strategy of early angiography in patients presenting with non-ST-segment elevation ACS. Whether this will be true as high-sensitivity troponin assays become available, particularly for levels below the 99th percentile of current assays, remains to be seen.

The role of high-sensitivity C-reactive protein (hsCRP) in guiding therapy with statins recently has come under scrutiny. Although the Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) showed that patients with normal low-density lipoprotein cholesterol (LDL-C) who had an hsCRP level > 2 mg/dl benefitted from treatment with rosuvastatin, there was no arm with similar LDL-C levels but hsCRP < 2 mg/dl. Thus, this trial cannot be used to argue that hsCRP should be used to guide statin treatment in primary prevention patients. Additionally, a recent cost-effectiveness analysis suggested that, assuming long-term safety of statins, availability of low-cost generic agents, and similar efficacy of statins in low-to-intermediate risk patients, treating men with statins without screening hsCRP would be cost-effective down to 50 years. At 70 years, using hsCRP to guide therapy would be cost-effective in both men and women; in both men and women, the lower bounds of age for primary prevention without hsCRP guidance rose with increasing numbers of cardiovascular risk factors.

As an example of the increasing interface of genetics with clinical care, there has been much interest in the use of genetic testing for the CYP2C19 mutation to guide clopidogrel therapy. However, despite associations of the mutation with outcome and demonstrated pharmacodynamic and pharmacokinetic variability with clopidogrel treatment according to carrier status, studies to date have not demonstrated that testing for this genetic mutation is useful in guiding treatment. A large randomized clinical trial, GRAVITAS, evaluated whether tailored clopidogrel dosing according to phenotypic platelet responsiveness measured prior to discharge after drug-eluting stent implantation would reduce thrombotic complications of stent implantation. Its results were presented at the 2010 Scientific Sessions of the American Heart Association (AHA) in Chicago and showed no benefit on cardiovascular outcomes or stent thrombosis with a double dose of clopidogrel in patients receiving drug-eluting stents with high residual platelet activity on the regular clopidogrel dose. These results are not yet published.

Thus, in 2010, it is increasingly evident that global risk assessment is needed to help clinicians align treatment with diagnosis and risk. Biomarkers play an important role in this process. However, rapid advances in assay technology and the increasing availability of new biomarkers generated from genomic discovery and applications of genetic testing create challenges that must be considered. Novel biomarkers must be systematic and rigorously evaluated, and their practical clinical utility must be demonstrated before they become part of a routine risk assessment strategy.

Measuring quality in ACS: where does antithrombotic therapy fit?

Quality of care has been defined as the "degree to which health care services increase the likelihood of desired health outcomes and are consistent with current professional knowledge." Simply put, this asks: Are we doing the right things (practicing evidence-based care); are we doing the right things right (delivering this care in a safe, skilled manner); and are our patients better off for it (are their outcomes improved)? When viewed in this manner, contemporary treatment of patients with ACS is challenged. Studies have consistently demonstrated an under-utilization of evidence-based therapies, as well as failure to provide such care in a safe and timely fashion. And while care is improving over time, consistent gaps remain. For example, 2010 data from the ACTION Registry®-Get With the Guidelines (GWTG)™ found that between 15 and 20% of eligible ACS patients fail to receive dual antiplatelet therapy acutely and at hospital discharge.

The standard application of evidence-based therapies, however, neglects to consider that these treatments ideally should be "personalized" for the individual patient. Antithrombotic therapies in ACS care effectively prevent recurrent ischemic events or, alternatively, cause iatrogenic bleeding. The balance between the benefits and risks is influenced by three domains. The first domain relates to features of the drug itself, including drug absorption, activation, potency, clearance, and interaction with other drugs. Patient factors represent a second domain influencing safety and efficacy of antithrombotic therapies in ACS, including such factors as patient age, sex, renal function, and presence of diabetes. These clinical features influence the baseline odds for recurrent ischemic events but also can affect the safety of antithrombotic therapy, either through

changing the drug's pharmacokinetic and dynamic properties or increasing the patient's underlying disposition to bleed (i.e., peptic ulcer disease).

Provider and system factors also influence the quality of care and subsequent outcomes in ACS. Studies have found that a number of patients in the United States receive the wrong dose of antithrombotic therapies. Combined, up to 20% of all bleeds in the United States are estimated to be caused by excessive antithrombotic therapies. The reasons for excessive drug dosing often relate to a failure to individualize dose based on body weight, age, or renal function.

While there are challenges to the effective and safe use of antithrombotic therapies in ACS, the world is changing, and efforts to improve the quality of ACS care delivered around the world abound. In particular, giving clinicians feedback on their care practices relative to those of their peers has been shown consistently to improve ACS quality of care. Moving forward, this follow-up and feedback regarding ACS practices must extend to consider longitudinal care and outcomes. For example, studies have consistently demonstrated that patients who discontinue dual antiplatelet therapy early after receiving a stent are at high risk for subsequent cardiac events. Importantly, patient compliance appears modifiable via patient education. Those who understand the reasons for their medications and the need for continued use have higher rates of compliance.

In the future, both providers and patients will have increasing access to electronic tools to facilitate better ACS care. These include electronic order entry systems that will support wiser drug choices and prevent medical errors related to drug dosing. We will also see the evolution of community systems of care that will encourage appropriate triage of ACS patients to support more timely ACS care. Finally, we will see the evolution of patient health records that will support a new collaborative model of care between patients and their caregivers.

Antiphospholipid syndrome

Antiphospholipid syndrome is widely recognized but incompletely understood. There are five areas worthy of consideration: pathophysiology, epidemiology, clinical manifestations, diagnosis, and management. The pathophysiology of antiphospholipid syndrome involves production of IgG antibodies against beta 2-glycoprotein I on the surface of vascular endothelial cells. The antibodies cause expression of adhesion molecules and up-regulation of tissue factor production. In addition, they produce up-regulation of tissue factor within monocytes, expression of GP IIb/IIIa receptors on platelets, and increased thromboxane A2 synthesis. The interaction of antibodies with coagulation regulatory proteins such as activated protein C in combination with complement activation and inflammation establishes a highly prothrombotic state. The available evidence suggests that an existing thrombophilia in antiphospholipid syndrome can be exaggerated acutely as part of a putative "second hit" phenomenon following trauma, infection, and other conditions in which a prothrombotic environment rapidly develops.

Antiphospholipid antibodies are detected in 20% of patients with an ischemic stroke before age 50 years, 20% of patients with VTE, 10–15% of women with recurring miscarriages, and 20% of women with a diagnosis of preeclampsia. The most common clinical manifestations of antiphospholipid syndrome, occurring in >20% of individuals, include VTE, thrombocytopenia, miscarriage or fetal loss, ischemic stroke or transient ischemic attack, migraine headache, and livedo reticularis. Less common clinical manifestations, occurring in 10–20% of individuals, include heart valve abnormalities, hemolytic anemia, and accelerated CAD. Unusual clinical manifestations, occurring in <10% of individuals, include seizures, vascular dementia, retinal artery or vein thrombosis, pulmonary hypertension, skin ulcers with digital gangrene, osteonecrosis, renal insufficiency, and mesenteric ischemia.

The diagnosis of antiphospholipid syndrome is supported by clinical criteria, including vascular thrombosis involving one or more episodes of arterial, venous, or small-vessel thrombosis in any tissue or organ. Thrombosis should be present without substantial evidence of inflammation within the vessel wall. A diagnosis of antiphospholipid syndrome in the context of pregnancy is supported by at least one of the following criteria: one or more unexplained deaths of a morphologically healthy fetus at or beyond the 10th week of gestation; one or more premature births of a morphologically healthy new born before the 34th week of gestation either because of eclampsia or severe preeclampsia; or at least three unexplained consecutive spontaneous abortions before the 10th week of gestation, with anatomical or chromosomal abnormalities having been excluded.

The laboratory diagnosis of antiphospholipid syndrome includes the following: detection of a lupus anticoagulant on two or more occasions at least 12 weeks apart; anticardiolipin antibody of IgG or IgM subtype, or both in serum or plasma, present in medium or high titers on at least two or more occasions at least 12 weeks apart measured with a standardized ELISA; or anti-beta 2 GP1 antibody of IgG or IgM subtype, or both in serum or plasma, at medium or high titers on at least two or more occasions at least 12 weeks apart.

The management of patients with antiphospholipid syndrome includes a strategy of primary prophylaxis,

where patients with systemic lupus erythematosus and a circulating lupus anticoagulant or persistently positive anticardiolipin antibody titer would receive hydroxychloroquine either alone or in combination with low-dose aspirin. Patients with obstetric antiphospholipid syndrome are traditionally treated with low-dose aspirin, while asymptomatic carriers of antiphospholipid antibodies do not typically require therapy. However, it is important to emphasize that all patients with antiphospholipid antibodies likely benefit from strict control of vascular risk factors and should receive adequate thromboprophylaxis in high-risk situations such as surgery, the post-partum period, and during prolonged periods of immobilization. Management of patients with antiphospholipid syndrome without previous thrombosis but recurring early (pre-embryonic or embryonic) miscarriages should include either low-dose aspirin alone or in combination with either UFH or LMWH. Patients with antiphospholipid syndrome without previous thrombosis but with prior fetal death at more than 10 weeks gestation or early delivery (<34 weeks gestation) due to severe preeclampsia or placental insufficiency should be treated with UFH or LMWH throughout pregnancy.

Secondary prophylaxis of patients with antiphospholipid syndrome and prior thrombosis typically includes indefinite anticoagulation with warfarin, titrated to a target INR of 2.5 (range 2.0–3.0). There is a suggestion that patients with a prior arterial thrombotic event should be targeted to a higher INR (3.5; range 3.0–4.0) or warfarin titrated to a target INR of 2.5 plus low-dose aspirin. The latter two strategies have also been used for patients with recurring events despite warfarin anticoagulation.

The potential use of newer-generation anticoagulants, such as oral direct factor Xa or direct thrombin inhibitors, will require further evaluation.

Venous thromboembolism prophylaxis in surgical patients

Venous thromboembolism is the most preventable cause of morbidity and mortality in postoperative settings; it is the second most common medical complication, the third most common source of excess health care resource utilization, and the third most common cause of mortality in postoperative patients. Accordingly, pulmonary embolism is the most common yet preventable cause of death.

Proximal vein VTE presents the highest risk for PE: 50% asymptomatic or ''silent'' PE and 25% of distal vein VTE will extend to proximal veins within 1 week of presentation. Most postoperative cases of VTE are clinically silent: 2–30% of in-hospital postoperative deaths are attributable to PE.

Over 30 million operations are performed annually in the United States, and the incidence of postoperative VTE without prophylaxis is 10–20% for low-risk and up to 80% in high-risk patients. The rates of fatal PE in the highest-risk patients range from 0.5 to 30%, with length of hospital stay of 5.4 days, excess mortality of 6.6%, and costs reaching $25,000 more than compared with controls.

There were approximately 38 million discharges in the United States in 2006: 7 million were surgical inpatients. According to American College of Chest Physicians (ACCP) Antithrombotic and Thrombolytic Guidelines risk categories, 44% of these patients were at low risk for VTE; 15, 24, and 17% were at moderate, high, and very high risk, respectively. Risk assessment strategy and systematic computerized electronic alerts should be a combined objective to increase the use of VTE prophylaxis and to reduce the rates of symptomatic VTE among hospitalized patients.

For prevention of this common problem, new evidence is available to support novel anticoagulant therapy.

Rivaroxaban

Evidence for this oral, direct factor Xa inhibitor for thromboprophylaxis was presented in the results from the Regulation of Coagulation in Orthopedic Surgery to Prevent Deep Venous Thrombosis and Pulmonary Embolism (RECORD) 1, 2, 3, and 4 trials (Table 1). The RECORD 1 trial was designed to evaluate oral rivaroxaban compared with subcutaneous enoxaparin for extended thromboprophylaxis after total hip arthroplasty. The primary outcome was total VTE: any deep vein thrombosis (DVT), non-fatal PE, and all-cause mortality at 36 days (range 30–42); secondary outcomes included major VTE: proximal DVT, non-fatal PE, and VTE-related death. DVT included any, proximal, distal, and symptomatic VTE.

In the RECORD 2 trial, extended thromboprophylaxis with oral rivaroxaban versus short-term subcutaneous enoxaparin following total hip replacement was evaluated. The main study question was whether extended-duration prophylaxis was superior to short-duration prophylaxis. In summary, RECORD 2 showed that extended-duration prophylaxis was superior to short-duration prophylaxis and that rivaroxaban provided an effective option for such a strategy and had a good safety profile.

Finally, RECORD 3 and 4 evaluated thromboprophylaxis after total knee arthroplasty and found that rivaroxaban (10 mg once a day for 10–14 days), given in a fixed, once-daily dose regimen without coagulation monitoring, was superior to enoxaparin (40 mg once a day for 10–14 days) in preventing venous thrombosis with similar rates of bleeding.

Table 1 Main efficacy outcomes of the RECORD trials

Trial	Setting	Enoxaparin regimen	Rivaroxaban regimen	DVT/PE/death (%)	RRR (%)	Symptomatic VTE (%)	RRR (%)
RECORD 1, $n = 4{,}541$	Hip arthroplasty	40 mg qd, 35 d	10 mg qd, 35 d	3.7 vs. 1.1	70	0.5 vs. 0.3	NS
RECORD 2, $n = 2{,}509$	Hip arthroplasty	40 mg qd, 10–14 d	10 mg qd, 31–39 d	9.3 vs. 2.0	79	1.2 vs. 0.2	80
RECORD 3, $n = 2{,}531$	Knee arthroplasty	40 mg qd, 10–14 d	10 mg qd, 10–14 d	18.9 vs. 9.6	49	2.0 vs. 0.7	66
RECORD 4, $n = 3{,}148$	Knee arthroplasty	30 mg bid, 10–14 d	10 mg qd, 10–14 d	10.1 vs. 6.9	32	1.2 vs. 0.7	NS

Bid twice daily; *d* days; *NS* not significant; *qd* daily; *RRR* relative risk reduction

Apixaban

Apixaban, an oral, direct factor Xa inhibitor, was evaluated for DVT prophylaxis after total knee replacement in a phase II dose-ranging study. Aggregated apixaban doses resulted in a 21% ($P < 0.02$) reduction in VTE and all-cause death compared with enoxaparin and a 53% ($P < 0.01$) reduction compared with warfarin. Major bleeding event rates were low (0–3.3%) and comparable across all apixaban arms and the enoxaparin and warfarin groups. Similar results were shown in a dose-ranging trial for the treatment of DVT.

The phase III, randomized, double-blind Apixaban Dose Orally versus Anticoagulation with Enoxaparin (ADVANCE)-1 trial compared the efficacy and safety of 2.5 mg oral apixaban twice daily to subcutaneous 30 mg enoxaparin for the prevention of VTE after total knee replacement in 3195 patients. The primary outcome rates in each arm were similar (8.99% vs. 8.85%). The predetermined non-inferiority end point was not met, but event rates were comparable, and there was less clinically relevant bleeding in the apixaban arm. There was no difference between the two groups in serious adverse events.

The Apixaban Dose Orally versus Anticoagulation with Enoxaparin (ADVANCE)-2 trial compared apixaban (2.5 mg orally twice daily) with enoxaparin (40 mg subcutaneously daily) for preventing VTE after total knee replacement. The primary efficacy outcome (all VTE) occurred in 15.1% of patients in the apixaban group and 24.4% in the enoxaparin group. A nonsignificant trend toward less clinically relevant bleeding also favored apixaban (3.5 vs. 4.8%, $P = 0.09$).

The Apixaban Dose Orally versus Anticoagulation with Enoxaparin (ADVANCE)-3 trial evaluated the efficacy and safety of oral, twice-daily apixaban 2.5 mg compared with subcutaneous enoxaparin 40 mg once daily in patients undergoing elective total hip replacement surgery. In this study, the primary efficacy end point occurred in 1.4% of patients in the apixaban group and 3.9% of patients in the enoxaparin group, demonstrating a statistically significant relative risk reduction for apixaban of 64% ($P < 0.001$ for non-inferiority and superiority). The safety outcome of major bleeding occurred in 0.8% of patients who received apixaban and in 0.7% of patients who received enoxaparin ($P = 0.54$). There was no difference between the two groups in serious adverse events.

Dabigatran

In patients undergoing total hip replacement who were enrolled in the Dabigatran Etexilate Compared with Enoxaparin in Prevention of VTE Following Total Hip Arthroplasty (RE-NOVATE) trial, both doses of the oral direct thrombin inhibitor, dabigatran etexilate, given for a median of 33 days were as effective as enoxaparin for the prevention of VTE, with a similar safety profile. Furthermore, dabigatran etexilate proved to be non-inferior to enoxaparin, when administered for the same duration, for reducing the risk of total VTE and all-cause mortality after total hip replacement. In patients undergoing total knee replacement who were enrolled in the Thromboembolism Prevention after Knee Surgery (RE-MODEL) trial, both doses of the oral direct thrombin inhibitor dabigatran etexilate, given for 6–10 days, were as effective as enoxaparin for the prevention of VTE, with a similar safety profile. Dabigatran etexilate proved to be non-inferior to enoxaparin (40 mg daily started the night before surgery) for the prevention of VTE after total knee replacement.

Should patients with cancer receive primary VTE prophylaxis?

Patients with cancer are at high risk for VTE, which is the cause of death in many patients with advanced malignancy. The risk for developing VTE is highest during the first 3 months after diagnosis and depends on many factors, including the use (and type) of chemotherapy as well as the

site and stage of the neoplasm. Recent randomized trials have confirmed the hypothesis that anticoagulants, especially when administered in therapeutic doses, can reduce the risk of VTE in at-risk cancer patients. Unfortunately, the absolute risk reductions achieved in several of the trials reported to date have been small and do not justify the hazards and costs associated with a strategy of routine prophylaxis in all patients. For example, the Prophylaxis of Thromboembolic Events in Cancer Patients Receiving Chemotherapy (PROTECHT) study enrolled 1,150 patients with advanced lung, breast, or colon cancer (all were receiving chemotherapy) and randomly assigned them to either nadroparin (prophylactic dose) or placebo. Although the proportion of patients experiencing the primary end point was lower in the treatment group (2.1 vs. 3.9%, $P = 0.033$), this small risk difference has not resulted in the adoption of primary prevention strategies for these populations. The PROTECHT results indicate that 55 patients would have to be treated with LMWH for 1 year to prevent one thromboembolic event.

At least two trials that have recruited patients with pancreatic cancer receiving gemcitabine (and that compared therapeutic-dose LMWH to placebo) have demonstrated more dramatic risk reductions for VTE but did not clearly show a survival advantage and have yet to be published in full manuscript form. Several groups have now validated the risk prediction model of Khorana et al.—a scoring system that has demonstrated that the risk of developing VTE increases with a number of factors, such as elevated white blood cell or platelet count, increased body mass index, or decreased hemoglobin. However, the absolute VTE risk level at which practicing oncologists and their patients should consider primary prevention remains unclear and may change if/when oral anticoagulants are shown to be effective for this purpose. At this time, the National Cancer Center Network guidelines do not recommend routine use of primary VTE prophylaxis in any outpatient cancer population, except for patients with multiple myeloma who are receiving lenalidomide and dexamethasone.

Are there patients with PE who can be treated out of hospital?

Pulmonary embolism is a common condition affecting more than 1.5 million Americans yearly. It is a serious disease that accounts for 10% of all in-hospital deaths and is a major contributing factor in another 10% of deaths. Despite these elevated death rates, PE might be a more benign condition when associated with a lower thrombus burden. In this case, mortality is extremely low, and patients might be considered for outpatient therapy. Therefore, risk stratification is of utmost importance when considering therapy in PE.

Several risk stratification scores have been developed and include variables such as age, clinical status at hospital admission (heart rate, blood pressure, and respiratory rate), presence of cancer, and hypoxemia ($SatO_2 < 90\%$). The commonly used Geneva risk score demonstrates good discrimination for the prediction of death, major bleeding, and recurrent VTE at 3 months. Patients stratified as low risk (80% of total) have a 2.2% event rate, whereas high-risk patients have increased risk of complications (26%). Another famous score called PESI (Pulmonary Embolism Severity Index) performs similarly for the prediction of the same end points. More recently, echocardiographic data and biomarker measurements, such as cardiac troponins (cTnT and cTnI) and brain natriuretic peptide (BNP), have been included in these scores. Biomarkers predict death and other complications following PE with an odds ratio as high as 17.9 according to some studies. They also improve discrimination beyond clinical and echocardiographic variables. In conclusion, patients admitted with PE are at different risk. Currently available risk stratification scores help predict complications and enable the choice of the most suitable therapy for each patient. A study from the Netherlands presented as a late-breaking session at the American Society of Hematology annual meeting in December of 2010 indicates that out-of-hospital therapy may be reasonable in selected patients with PE.

Approaches for patients with venous thrombosis in unusual sites

The vast majority of proven episodes of DVT occur in the deep veins of the legs. When they occur in the proximal veins, embolization may travel to the lungs, producing PE. DVT and PE are often described as VTE and comprise a leading cause of hospital-acquired morbidity and mortality.

Venous thrombosis may occur in any vein. Recently, the frequency of DVT in non-leg veins has increased dramatically due to the increased sensitivity of our radiologic investigations. For example, improved resolution of abdominal ultrasonography and contrast-enhanced computed tomography (CT) scanning has led to a rapid increase in the frequency of detection of splanchnic venous thrombosis, oftentimes occurring in patients with minimal or no referable symptoms who are undergoing evaluation for unrelated medical indications. Cerebral vein thrombosis is a potentially devastating form of thromboembolism that is optimally detected with magnetic resonance venography or direct angiography. Again, due to the increasing availability and resolution of these modalities, the frequency of detection of these thrombi is increasing. Finally, DVT may

occur in other vascular sites such as the renal veins, pelvic veins, pulmonary veins, and in varicose veins located in any vascular distribution.

Thrombophilia testing is widely available and grossly overused. However, there appears to be a particular predilection for patients with selected forms of thrombophilia to develop thrombosis in unusual sites. For example, patients with paroxysmal nocturnal hemoglobinuria (PNH) appear to be particularly prone to develop Budd-Chiari syndrome, while patients with the JAK-2 mutation appear prone to splanchnic vein thrombosis. A recent systematic review demonstrated that almost one third of patients presenting with splanchnic venous thrombosis had the JAK-2 mutation, with many patients having a normal complete blood cell count. The JAK-2 mutation (although recently discovered) has traditionally been identified as being characteristic of myeloproliferative disorders. The mechanism by which this mutation predisposes a patient to splanchnic venous thrombosis is unknown. Both the lupus anticoagulant and anticardiolipin antibody are frequently detected in patients presenting with unusual forms of thrombosis, particularly at young ages. Detection of antiphospholipid antibodies, including both lupus anticoagulant and anticardiolipin antibody, is important because most experts would recommend extended-duration therapy with an oral anticoagulant in patients with these antibodies. There are no specific therapies for the JAK-2 mutation, and patients appear to be treated effectively with oral anticoagulants. Patients with PNH may be resistant to warfarin administered to a traditional INR between 2 and 3; recent studies have suggested a high rate of "warfarin failure." Eculizumab is a recently approved medication that blocks the terminal complement components and thus reduces hemolysis in patients with PNH. Indirect evidence suggests that this medication may also ameliorate the thrombotic complications of this disorder.

More common thrombophilias, such as the prothrombin gene mutation, appear to be particularly common in patients with cerebral vein thrombosis.

Ovarian vein and other pelvic vein thrombosis appear to be particularly common in the peripartum period. Renal vein thrombosis is particularly common in patients with renal cell carcinoma and may be more common in patients with nephrotic syndrome. DVT of the upper extremity is particularly common in the setting of indwelling central venous catheters and in athletes, presumably because of impingement on the veins leaving the arm during vigorous exercise.

Patients with unusual site thrombosis appear to respond to anticoagulation, with similar recurrence rates as patients with PE or thrombosis in the deep veins of the leg. Thus, a rapid-acting parenteral anticoagulant should be administered initially and overlapped with an oral VKA. This therapy may need to be modified as a result of studies of novel agents that may or may not require the initial course of parental anticoagulants.

There is no evidence as to how patients with "asymptomatic" clots should be treated. Most experts would treat patients with thrombi discovered in the setting of cancer or other high-risk situations. If there is reasonable evidence of prior thrombosis, then it may be reasonable to not anticoagulate. Patients who appear to be at high risk of complications, however, probably should be anticoagulated using a rapid-acting, parenteral anticoagulant overlapped with an oral VKA.

The duration of therapeutic anticoagulation has not been studied in these patients. Most experts extend anticoagulation because of a perception that recurrent disease could be associated with catastrophic complications. However, there is reasonable evidence that anticoagulants can be safely discontinued in selected patients, particularly in those with cerebral vein thrombosis.

In summary, DVT may occur in any vein. Thrombophilias appear to be particularly common in patients with unusual site thrombosis. Recent attention has focused on the JAK-2 mutation and PNH as causes of splanchnic and hepatic vein thrombosis, respectively. Anticoagulant therapy is indicated for all symptomatic patients. Optimal therapy of patients with screening-detected clots is unknown. In general, anticoagulant therapy is extended in patients with unusual site thrombosis due to the potentially catastrophic implications of recurrence.

Debate: VTE prophylaxis should be the default position for hospitalized medical patients—for/against

Thromboprophylaxis: the case against

There is no question that selected patients admitted to the hospital with medical disorders are at high risk of DVT and PE, oftentimes described as VTE. However, recommendations for the use of VTE prophylaxis have tended to err on the side of suggesting prophylaxis for most patients, despite a singular lack of evidence to support this recommendation. The ACCP guidelines recommend strongly that anticoagulant prophylaxis be provided to patients identified at high risk of VTE. Such patients include those with an extended duration of immobilization, congestive heart failure, serious thrombophilias, or those with more than one risk factor.

The case that VTE prophylaxis is not required in all patients is made simply. Patients with active bleeding or those perceived to be at very high risk of bleeding should not receive pharmacologic prophylaxis, thus establishing that there is a small but important subgroup of patients in whom prophylaxis is contraindicated. The bigger question

is whether VTE prophylaxis should be provided to low-to-moderate-risk patients.

Prophylaxis may be mechanical or pharmacologic. Mechanical prophylaxis can be active or passive. Passive mechanical prophylaxis, most commonly manifest as graduated compression stockings (GCS), are probably effective for the prevention of VTE but are significantly expensive when routinely used across a hospital and may be associated with transmission of infection. Intermittent pneumatic compression devices are probably more effective than passive compression devices, but they are expensive and are generally poorly used in hospitalized patients. Furthermore, reuse of intermittent compression device bladders may be associated with infectious diseases such as methicillin-resistant *Staphylococcus aureus*. There is no high-quality evidence by which to gauge the effectiveness of intermittent pneumatic compression devices in medically ill patients.

Pharmacologic prophylaxis generally consists of heparin or LMWH administered twice or three times daily. There is clear evidence that, in high-risk patients, pharmacologic prophylaxis reduces the risk of symptomatic DVT, PE, and fatal PE. However, there have been no studies demonstrating the effectiveness of these agents in the low-to-moderate-risk patient.

Irrespective of the indication, there is clear evidence that anticoagulants administered at prophylactic doses increase the risk of major bleeding. Major bleeding is expensive to treat and may be fatal in rare cases.

Modeling of the impact of prophylaxis provision on low-to-moderate-risk patients suggests that the risk of PE and fatal PE is low and very low, respectively. Although it is logical to assume that pharmacologic prophylaxis would further reduce the risk of thrombosis, there is also little doubt that prophylaxis would increase the risk of major and fatal bleeding. Rough modeling suggests, in fact, that, in low-risk patients, provision of prophylaxis would actually cause more fatal bleeding episodes than it prevented through reduced risk of PE. Additionally, the routine use of prophylaxis increases direct drug acquisition costs and dramatically increases the costs associated with the management of bleeding complications.

Based on the lack of evidence of efficacy, indirect but highly suggestive evidence of toxicity, a likely adverse cost-effectiveness profile, and the possibility that prophylaxis delivered to low-to-moderate-risk patients may actually increase the risk of death, it is clear that it is inappropriate to recommend VTE prophylaxis uniformly for medical patients.

Thromboprophylaxis: the case for

Pulmonary embolism is among the leading causes of death among patients hospitalized for acute medical illness. Although effective mechanical and pharmacologic modalities are available to reduce the risk of PE and DVT, clinicians often do not employ VTE prevention strategies for at-risk patients admitted with nonsurgical illnesses. The reasons for this underutilization of effective prophylaxis are not known with certainty; however, there are probably many factors involved. First, the physician caring for a patient with acute medical illness can easily be distracted by many other demands for his/her attention; VTE prevention can easily be forgotten. Second, a validated, user-friendly scheme by which medical patients can be stratified according to VTE risk does not exist. In light of the potential for "sensory overload" among inpatient physicians and the lack of an easy-to-use risk assessment model, it is unreasonable to expect health care providers reliably to prescribe prophylaxis against VTE to at-risk patients.

Cost is not a reason to oppose the routine use of VTE prevention strategies among medical patients. UFH and graduated compression stockings are relatively inexpensive, and there is high-quality evidence that both will reduce the risk of symptomatic DVT and PE. The absolute risk of major bleeding is not substantially increased by the use of low-dose anticoagulants (e.g., LMWH, fondaparinux, UFH) in this population. In other words, because of low baseline risk of bleeding in this population, it is likely that well over 100 patients would have to be treated with low-dose anticoagulants (versus nothing) to cause one additional major hemorrhage. While it would certainly be reasonable to withhold anticoagulants from a patient at high risk for bleeding (e.g., a cancer patient with profound thrombocytopenia), the "default" position should be to provide VTE prophylaxis to all medical patients because: 1) even patients at high risk for bleeding can benefit from mechanical interventions, and 2) for the vast majority of patients at "average" risk for bleeding, the trade-off will favor low-dose anticoagulants.

Triple therapy: patients with CAD and AF

Patients with cardiovascular disease may have several concomitant indications for antithrombotic therapy including ACS, DVT and PE, mechanical valves, AF, and coronary stent implantation. Overlapping indications for antithrombotic therapy may lead to the need for "triple therapy," defined here as aspirin, clopidogrel, and oral anticoagulation.

As the population ages, more patients will have both ACS and AF; accordingly triple therapy may be used more frequently. Prior studies have shown that, with more antithrombotic therapy, risk of bleeding increases. Many antiplatelet and anticoagulant drugs are part of the foundation for treatment of ACS and AF, making the decision

about the right combination of these agents challenging. However, limited evidence is available to guide therapeutic decision-making about triple therapy. Registry information, subgroup analyses from clinical trials, and overviews of single-center experiences have been published, but no randomized trials evaluating different strategies of triple therapy have been completed.

Multiple guidelines and consensus statements from national societies provide recommendations for clinicians concerning the use of triple therapy. A simple flow diagram can be used by physicians to guide decisions about the need for dual antiplatelet therapy or triple therapy based on the assessment of patient bleeding and stroke risk. Five additional factors should be considered: 1) use of the lowest dose of antiplatelet therapy; 2) use of bare metal stents versus drug-eluting stents to minimize the duration of antiplatelet therapy; 3) optimal INR within a range of 2.0–2.5; 4) gastric protection with PPIs; and 5) minimization of the duration of triple therapy. It is also important to re-evaluate regularly the need for triple therapy. The risk of stent thrombosis will decrease over time, whereas bleeding risk will remain constant.

Two ongoing randomized clinical trials will evaluate the role of triple therapy: the What Is the Optimal Antiplatelet and Anticoagulant Therapy in Patients with Oral Anticoagulation and Coronary Stenting (WOEST) study of $\sim$ 500 patients post-stenting randomized to triple therapy versus dual therapy (clopidogrel and an oral anticoagulant) and the Intracoronary Stenting and Antithrombotic Regimen: Testing of a Six-Week Versus a Six-Month Clopidogrel Treatment Regimen in Patients With Concomitant Aspirin and Oral Anticoagulant Therapy Following Drug-Eluting Stenting (ISAR-Triple) trial of $\sim$600 patients post-drug-eluting stent implantation randomized to triple therapy for 6 weeks versus triple therapy for 6 months.

Several new antiplatelet and anticoagulant agents are also being studied for ACS and AF, including the PAR-1 inhibitors in the Thrombin Receptor Antagonist for Clinical Events Reduction (TRACER) and TRA-2P programs; factor Xa inhibitors in the Anti-Xa Therapy to Lower Cardiovascular Events in Addition to Aspirin with or without Thienopyridine Therapy in Subjects with Acute Coronary Syndrome-Thrombolysis In Myocardial Infarction 46 (ATLAS ACS-TIMI 46) and the Apixaban for Prevention of Acute Ischemic Events (APPRAISE)-2 ACS trials; and factor Xa inhibitors in the Global Study to Assess the Safety and Effectiveness of DU-176b versus Standard Practice of Dosing with Warfarin in Patients with Atrial Fibrillation (ENGAGE AF-TIMI 48), Apixaban for the Prevention of Stroke in Subjects with Atrial Fibrillation (ARISTOTLE), and the Efficacy and Safety Study of Rivaroxaban with Warfarin for the Prevention of Stroke and Non-Central Nervous System Systemic Embolism in Patients with Non-Valvular Atrial Fibrillation (ROCKET AF) trials. The future will be interesting. Triple therapy may actually be redefined in the future with new $P2Y_{12}$ inhibitors such as prasugrel and ticagrelor, oral factor Xa inhibitors, and antithrombin agents. Indeed, warfarin may become obsolete in patients with ACS and AF. In addition, triple therapy may be replaced by "quadruple therapy" with aspirin, the $P2Y_{12}$ inhibition, PAR-1 inhibition, and oral anticoagulants.

Measuring quality in atrial fibrillation

Atrial fibrillation is a major health concern as assessed by almost any metric. Over 3 million U.S. citizens have AF, a number that is expected to nearly double by the year 2050. Patients with either paroxysmal or persistent AF have three-to-five-fold increased risk for stroke, and AF accounts for up to 75,000 strokes per year (15% of all U.S. strokes). Furthermore, those with AF have significantly higher mortality and lower quality of life than those without.

Treatment of AF is complex but centers on two major goals: reducing patients' embolic risk and controlling their symptoms. Oral anticoagulant therapies (e.g., VKAs like warfarin) are extremely effective in reducing patients' risk for stroke. However, the use of warfarin is complex and concomitantly can increase patients' risk for bleeding events. Thus, warfarin use is reserved for those with at least moderate stroke risk.

Current American College of Cardiology/American Heart Association (ACC/AHA) AF performance indicators include assessing the thromboembolic risk (CHADS score), initiating warfarin in those with moderate or high risk, and then closely monitoring warfarin therapy to ensure that patients are in a narrow therapeutic range. Opportunities for improvement on each of these performance metrics abound. Depending on the study, only about 30–60% of eligible AF patients in community practice actually receive warfarin therapy. Those at highest risk, as assessed by the CHADS score, are paradoxically less likely to receive warfarin therapy. And even when instituted, time-in-therapeutic range (TTR)—an important indicator of warfarin's safety and effectiveness—ranges from 30 to 60% in community case series. Newer agents, such as the oral direct thrombin and factor Xa inhibitors, represent a major leap forward for antithrombotic therapy for AF. These new drug classes offer easier patient management without constant drug monitoring. Furthermore, relative to warfarin, these new drugs are being demonstrated to have similar or improved thrombotic protection and significantly better safety profiles.

The second goal of AF management is to control patient symptoms and improve quality of life. Rhythm control of AF, either with anti-arrhythmic drugs or with AF ablation procedures, can restore sinus rhythm in many patients with AF. However, studies to date in mildly symptomatic patient subgroups have had difficulty showing that restoration of sinus rhythm necessarily improves quality of life or reduces stroke risks; further research is needed.

New anticoagulants for stroke prevention in atrial fibrillation

Warfarin is effective for stroke prevention in AF but has limitations because of variability in response and an increased risk of bleeding. The most feared complication of warfarin is intracranial bleeding. The efficacy and safety of warfarin is related to the TTR, which is an INR of 2.0–3.0; there is an increased risk of stroke and death at INR < 2 and of bleeding at INR > 3. However, the risk of bleeding, including intracranial bleeding, is present also in patients within the target range. This limits the indication for warfarin to patients with an intermediate-to-high risk of stroke (i.e., with a $CHADS_2$ risk score above 1) to maintain the net clinical benefit.

Therefore, development of new oral anticoagulants aims to demonstrate that they are at least as effective as warfarin and with better safety, allowing use in lower-risk populations. The new alternatives provide more specific inhibition of the coagulation cascade (i.e., by inhibition of thrombin [dabigatran] or factor Xa [apixaban, rivaroxaban, edoxaban, betrixaban]). Currently, the final results from prospective trials comparing these new treatment alternatives to warfarin in patients with AF and an increased risk of stroke are available for dabigatran from the pivotal Randomized Evaluation of Long-term Anticoagulant Therapy Warfarin Compared with Dabigatran (RE-LY) trial performed with a PROBE design. However, prospective double-blind trials comparing apixaban and rivaroxaban, respectively, with warfarin in similar populations have been presented or will be presented within the next year.

The ROCKET AF trial was presented at the AHA Scientific Sessions in November 2010. This study was a prospective, randomized, double-blind, double dummy, parallel-group, multicenter, event-driven non-inferiority study comparing the safety and efficacy of dose-adjusted warfarin with rivaroxaban 20 mg once daily. The primary efficacy end point for non-inferiority in ROCKET AF was the composite of stroke (ischemic and hemorrhagic) and non-central nervous system systemic embolism. The rate of primary outcome per 100 patient-years was 2.12 in the rivaroxaban arm compared with 2.42 in the warfarin arm ($P = 0.117$ for superiority, $P < 0.001$ for non-inferiority). Rivaroxaban also had a slightly better mortality profile: 582 deaths versus 632 in the warfarin group, but the difference was not statistically significant. In a per-protocol analysis, rivaroxaban was superior to warfarin with a primary outcome rate of 1.71 per 100 patient-years versus 2.16 ($P = 0.018$ for superiority and $P < 0.001$ for non-inferiority). Importantly, patients treated with rivaroxaban had fewer intracranial hemorrhages (0.49 vs. 0.74%, $P = 0.019$), fewer critical organ bleeds (0.82 vs. 1.18%, $P = 0.007$) and lower bleeding-related deaths (0.24 vs. 0.48%, $P = 0.003$) than those on warfarin. Rivaroxaban was well tolerated in the study, and rates of discontinuation due to adverse events were similar to those seen for patients on warfarin. One major criticism of the study was the poor INR control compared with previous AF trials. Among warfarin patients, the median time spent within therapeutic range was just 57.8%; they were above therapeutic range 11.9% of the time and below range 19.7% of the time. The results of the study are not published yet.

The RE-LY trial randomized 18,113 patients with AF in 951 sites to blinded fixed doses of dabigatran 110 mg or dabigatran 150 mg twice daily versus unblinded warfarin dose adjusted to INR 2.0–3.0. Median follow-up was 2 years. Rates of the primary outcome were 1.70% per year on warfarin versus 1.55% per year on dabigatran 110 mg (P non-inferiority < 0.001) and 1.11% per year on dabigatran 150 mg (P superiority < 0.001). Rates of major hemorrhage were 3.46% per year on warfarin versus 2.74% per year on dabigatran 110 mg ($P = 0.002$) and 3.22% per year on dabigatran 150 mg ($P = 0.32$). Rates of hemorrhagic stroke were 0.38% per year on warfarin versus 0.12% per year on dabigatran 110 mg ($P < 0.001$) and 0.10% per year on dabigatran 150 mg ($P < 0.001$). Mortality rates were 4.13% per year on warfarin versus 3.74% per year on dabigatran 110 mg ($P < 0.12$) and 3.63% per year on dabigatran 150 mg ($P < 0.047$).

Continued analyses of the RE-LY database have investigated the relative effects of dabigatran in relation to the average time in therapeutic range (cTTR) in each center's warfarin population and to $CHADS_2$ score. The quartiles of cTTR for the warfarin patients were <57, 57–65, 65–73, and >73%. There were no significant interactions with cTTR concerning the superiority of dabigatran 150 mg or the non-inferiority of dabigatran 110 mg versus warfarin for prevention of stroke and systemic embolism and both doses' superiority concerning intracranial bleeding. With dabigatran 150 mg, there was less major bleeding and lower but similar bleeding at higher quartiles of cTTR, while the rates of major bleeding were lower with dabigatran 110 mg irrespective of cTTR. Total mortality was lower with both dabigatran doses at lower cTTR levels and similar at higher cTTR levels.

In the RE-LY trial, around one third of patients had $CHADS_2$ scores 0–1, 2, or 3–6. Increasing $CHADS_2$ scores were associated with increased risks for stroke, bleeding, and mortality, with consistent benefits of dabigatran across all $CHADS_2$ risk groups above 0. Also, patients with the highest risk for new events (i.e., those with previous stroke) had consistent benefits with dabigatran versus warfarin.

Recently, the Apixaban versus Acetylsalicylic Acid to Prevent Strokes (AVERROES) trial compared the factor Xa inhibitor, apixaban (5 mg b.i.d.), with aspirin (80–325 mg/day) for stroke prevention in patients with AF who were unsuitable for oral anticoagulation. The trial was prematurely terminated because apixaban was found superior to aspirin in prevention of the primary end point of stroke and systemic embolism: there was a 54% reduction ($P < 0.001$) at a mean follow-up of 1.1 years. There was no significant difference in major bleeding or any other major safety end point. Apixaban was better tolerated than aspirin, with fewer discontinuations of apixaban compared with aspirin (RR 0.88, 95% CI 0.78–1.00, $P = 0.04$).

In summary, for patients with AF, direct thrombin inhibition with dabigatran provides an attractive alternative to warfarin therapy that preserves or improves on the reduction in stroke and systemic embolism achievable with warfarin with similar to lower rates of major hemorrhage. Factor Xa inhibition with apixaban offers a superior alternative to aspirin for stroke prevention in AF patients who are not candidates for warfarin, with even better tolerance than aspirin. The role of rivaroxaban or apixaban in treatment of warfarin-eligible patients awaits peer-reviewed data from ongoing or recently completed studies. Therefore, there is great hope that soon several new treatment alternatives will be available for stroke prevention in AF that should improve both patient outcomes and quality of life.

The relative importance of stroke and bleeding risk in patients with AF: a case-based approach

You are seeing a new patient in clinic. She is an 82-year-old female with hypertension, diastolic heart failure, and non-valvular AF. She has no idea how long she has been in AF, and she reports no change in her symptoms. Her heart rate is irregular, 85 beats per minute, and her blood pressure is 130/80 mmHg. She asks, "Should I start warfarin?" Initially, this seems like a relatively easy question; however, the decision to start a patient on life-long anticoagulation requires a careful assessment of benefits and risks of anticoagulation and consideration of how this information should be used for an individual patient.

Evidence-based medicine, as described by David Sackett, is the process of combining quantitative evidence about medical practice with expert physician judgment to ensure each individual patient the best medical care with reproducible high quality. To provide evidence-based thromboembolism prophylaxis in patients with AF, one has to carefully consider the benefits of thromboembolism prophylaxis (primarily a reduction in the risk of thromboembolic stroke) and the risks of thromboembolism prophylaxis (primarily an increase in the risk of bleeding). These population-based benefits and risks then need to be applied to the individual patient.

The absolute risk of stroke in patients with AF is less related to the burden of AF and more related to patient comorbidities. A number of risk scores have been developed. The most common is the $CHADS_2$ score, which assigns one point for heart failure, hypertension, age $>$ 75 years, and diabetes, and two points for prior stroke. The risk of stroke increases with increasing $CHADS_2$ score, from roughly 2% per year for $CHADS_2$ scores of 0–1 to over 15% per year for $CHADS_2$ scores of over 6. A newer score, the CHADS-VASC, includes points for female sex, vascular disease, and age between 65 and 75 years, and assigns two points for age $>$ 75 years. The CHADS-VASC score better stratifies risk in patients with a $CHADS_2$ score of 0. Our patient has a $CHADS_2$ score of 3 and a CHADS-VASC score of 4. Based on this, her annual risk of stroke is 6–8%. She says, "I'm old and understand I have a risk of stroke, but should I take warfarin?"

There are two additional important factors that have to be incorporated when considering the potential benefits of warfarin for this patient. The first is just how bad a stroke is likely to be and the second is whether warfarin will be effective at reducing the risk of stroke. The definition of stroke used in most of the clinical trials of thromboembolism prophylaxis in patients with AF is non-traumatic, focal neurologic deficit lasting at least 24 h. Thus, some strokes are devastating, while others result in no long-term deficit. However, strokes in patients with AF tend to be severe, with more than two-thirds resulting in death or permanent disability. Also important is that warfarin is highly effective at reducing strokes in patients with AF. Treatment with warfarin results in a roughly two-thirds reduction in stroke. Therefore, our patient has a more than 4% risk per year of a disabling stroke, and her risk of stroke could be reduced to roughly 2% with warfarin.

Warfarin, a potent anticoagulant, has bleeding as its major side effect. Warfarin is most effective in patients who maintain an INR between 2 and 3. With an INR below 2, the risk of stroke promptly increases. With an INR above 3, the risk of bleeding increases. However, even with reasonably good INR control, patients taking warfarin have a roughly 2% annual risk of major bleeding. The risk factors for bleeding substantially overlap with the risk factors for stroke. The recently developed

HAS-BLED score assigns one point each for hypertension, abnormal renal function, abnormal liver function, prior stroke, a history of bleeding, poor INR control, age > 65 years, and drug and alcohol use. The risk of bleeding ranges from 1% with a HAS-BLED score of 0 to more than 15% with a HAS-BLED score of 5. Our patient has a HAS-BLED score of 2 or more; thus, an annual risk of major bleeding on warfarin of 3% or more. She asks, "How bad is major bleeding?"

Like stroke, it is important to consider the range of major bleeding. The definition of major bleeding in most clinical trials of thromboembolism prophylaxis in patients with AF is that of the International Society of Thrombosis and Hemostasis (ISTH). ISTH major bleeding includes fatal bleeding, symptomatic intracranial, intra-articular, intra-spinal, pericardial, intraocular, retroperitoneal, or intramuscular bleeding with compartment syndrome, or bleeding resulting in a fall in hemoglobin of at least 2 g/dl or leading to transfusion of two or more units of red blood cells. In the recently reported RE-LY trial, the rate of major bleeding with warfarin was 3.6%, while the rate of the most devastating intracranial bleeding was only 0.7%. These were in contrast to a stroke rate of 1.6%. Intracranial bleeding is consistently associated with much worse outcomes than other types of major bleeding. When one considers a "net clinical benefit" that includes reduction in stroke and increase in only intracranial bleeding, those patients with a $CHADS_2$ score of 2 or more have a significant benefit with warfarin. This includes our patient above who, with a $CHADS_2$ score of 3, would be expected to have a net benefit of roughly 2% per year with warfarin. Now that we have covered the major efficacy and safety issues with warfarin, our patient asks, "Are there are any other downsides to warfarin?"

Warfarin, although one of the most effective drugs available to prevent devastating consequences of atrial fibrillation, also has significant downsides beyond bleeding. Warfarin has a host of dietary and drug interactions and requires at least monthly INR monitoring; many patients are plagued by significant INR variability requiring frequent dose changes. Finally, and perhaps most importantly, the dietary and drug interactions and the need for frequent monitoring create a constant worry on both the part of the patient and his or her physician. It is for these reasons that warfarin isn't used in close to half of patients with AF, including many of those who are at the highest risk of stroke. Fortunately, for all patients with AF, there are a host of alternatives to warfarin, including factor X and factor II (thrombin) inhibitors that are in development. Some of these may offer better efficacy and/or safety than warfarin, but all are likely to result in less worry; thus, hopefully, we will see more use of effective thromboembolism prophylaxis in patients with AF. Based on this discussion, our patient has decided to start warfarin as thromboembolism prophylaxis, at least until one of these alternative anticoagulants is available.

The anticoagulation of STEMI patients not eligible for reperfusion

In clinical practice, approximately 30% of patients with STEMI will not receive reperfusion therapy, either by primary PCI or lytics, because of delayed presentation, increased risk of bleeding, or patient-related factors. Systemic anticoagulants have been tested in this setting as a way to reduce the occurrence of adverse events, including mortality and re-infarction. A limited number of contemporary trials are available to guide clinical decision-making; however, there is no clear consensus on the use of systemic anticoagulation in this setting.

In a post-hoc analysis of the Thrombolysis in Myocardial Infarction (TIMI) 11B and Efficacy and Safety of Subcutaneous Enoxaparin in Non-Q wave Coronary Events (ESSENCE) trials, of the 7081 patients initially identified as having non-STEMI, 252 were subsequently found to have Q-wave MI. When treated with enoxaparin instead of UFH, these patients had 28% less death, MI, and recurrent angina at 30 days ($P = 0.04$). These results prompted the Treatment of Enoxaparin and Tirofiban in Acute Myocardial Infarction (TETAMI) trial, which, to this day, remains the only prospective randomized controlled trial specifically testing anticoagulation strategies in STEMI patients not eligible for early reperfusion therapy. TETAMI compared the efficacy and safety of enoxaparin versus UFH and eptifibatide versus placebo in a factorial design. In this context, enoxaparin and UFH were equivalent in terms death, re-infarction, or recurrent angina at 30 days (15.7 vs. 17.3%, respectively; OR = 0.89, 95% CI 0.66–1.21, P = NS). Tirofiban was not superior to placebo to improve outcome and tended to increase the rate of major bleeding.

In a more contemporary setting, the randomized double-blind OASIS-6 trial compared fondaparinux with UFH or placebo in STEMI patients, some of whom who were not eligible to receive reperfusion. In this subgroup, fondaparinux was better than either UFH or placebo at reducing the occurrence of death or MI at 30 days (12.2 vs. 15.1%, $P = 0.04$). Interestingly, the rate of major bleeding among patients treated with fondaparinux was similar to controls (hazards ratio = 0.84, 95% CI 0.47–1.50, $P = 0.55$).

Despite our best efforts, a significant proportion of patients do not receive reperfusion therapy. In 2010, we don't know with certainty that anticoagulation is superior to no anticoagulation in patients with STEMI not eligible for reperfusion.

Antiplatelet therapy in patients undergoing CABG surgery: what should we do?

Antiplatelet therapy significantly reduces mortality in ACS. However, a problem is posed when patients on antiplatelet therapy require cardiac surgery, as this represents a recognized risk for increased surgical blood loss. Formerly, aspirin was suspended for 5 days before surgery; in recent years, this approach has not been routinely followed. In fact, some centers introduce aspirin before surgery, especially when an off-pump technique is employed. The same practice of introducing aspirin before surgery does not occur with clopidogrel, which is most widely used in ACS after drug-eluting stent implantation and before primary PCI for MI. Clopidogrel is administered in emergency rooms when ACS is suspected, even before a definitive diagnosis is made. It permanently blocks platelets, and its effect only diminishes after the natural platelet replacement, which takes 5–7 days in a normal subject. CABG surgery should be avoided during this period, but this is not strictly observed in practice, nor is it clear the magnitude of the contribution of clopidogrel to surgical bleeding. Short-acting and reversible antiplatelet drugs, such as ticagrelor (oral) and cangrelor (intravenous), are being introduced, but they are not yet in general clinical practice. Abciximab, a humanized monoclonal antibody to the platelet GP IIb/IIIa receptor, irreversibly binds the receptor, has a more intensive antiplatelet effect, and should be avoided before surgery. However, the small-molecule GP IIb/IIIa inhibitors (eptifibatide and tirofiban) reversibly bind the receptor, have short half-lives, and have not been shown to increase CABG-related bleeding.

Most guidelines and practicing cardiac surgical centers recommend stopping clopidogrel administration for 5 days before elective surgery. In one multicenter analysis, exposure to clopidogrel within 5 days before CABG was associated with a 9.8-fold increase in need for reoperation ($P < 0.01$). However, other analyses have found weaker or no relationship with reoperation. In urgent situations, the risk of MI or its extension must be balanced against the risk of surgical bleeding, increased morbidity, and possible mortality. Intravenous UFH, on the other hand, could be safely and efficaciously introduced instead of clopidogrel in emergency situations, until a coronary artery anatomical diagnosis is obtained and a decision for PCI or CABG is made. One strategy commonly used is to not administer clopidogrel or prasugrel until the anatomy is known in STEMI patients. However, in NSTE ACS patients, the overall rate of CABG is only 10–15%, and there are no effective methods to predict at presentation who those individuals will be. Thus, the ischemic benefits of early treatment in this situation may outweigh the downsides of delay to CABG if it is ultimately indicated.

In the unstable patient with severe proximal coronary artery lesions, when surgery must be performed in the presence of clopidogrel, some adjuvant measures for better hemostasis may be considered, though few data are available for their effectiveness. These include careful surgical evaluation, the use of prophylactic antithrombotic agents, such as epsilon-aminocaproic acid or tranexamic acid (not aprotinin), during and after the procedure, and platelet infusion.

Statistical issues in the design and analysis of clinical research

As new drugs and devices are developed, questions arise as to the efficacy and safety of these treatments overall and relative to other available treatments, as well as to which patient populations would benefit most from the new therapy. The ideal situation for answering these questions would be to treat the entire population of eligible patients and observe all responses. But it is usually impossible to treat and evaluate every possible patient. Instead, we study the use of the therapy in a sample of the population. Based on the results observed in the sample, we make inferences about what we would expect to see if we could have applied the treatment to the entire population.

Multiple aspects of research determine the level of confidence one can have that the results observed in the sample are real and not just an anomaly of that sample or experiment. The number of patients studied must be large enough to provide adequate power to detect a significant difference. The patients studied should be generalizable to the population of interest. The allocation of treatment to the patients must be in a random fashion to ensure no biases are introduced during the selection process. The blinding of treatment is another important step in eliminating bias. When possible, the treating physician, the patient, and all others involved in the study should be blinded to the treatment that the patient is receiving.

All aspects of the study should be clearly specified and well-defined. When possible, the actual end point of interest should be studied rather than a surrogate end point. For the end point, the definition should be explicitly described, thus allowing for reproducibility in future studies. The protocol should state upfront whether the results will be based on the enrolling physician's determination, independent core laboratory results, or an independent adjudication committee determination of the end point. The timing of the end point should be based on clinical relevance. With long-term outcomes, the short-term results are also known. But the treatment may only affect outcomes acutely, so results may become diluted after an extended period of time.

Randomization of the treatment of interest is not always possible. There are situations in which randomization would be unethical and/or impossible to implement. In these cases, we must instead study series of patients and use special statistical tools to account for biases. These include adjusting for confounders or for the propensity to receive one treatment versus another. If the modeling process can fully adjust for all of the factors that are associated with receiving the treatment and with the outcome of interest, then one can make causal inferences. However, this situation is seldom possible.

With multiple treatments for the same condition, the growing increase of genetic markers, globalization of clinical trials, and many changes in research over the past few years, the analytic issues have become increasingly complex. Statistical expertise is needed to ensure high-quality, accurate results. A greater understanding of the underlying statistical issues in clinical research is needed for the non-statistician, who must critically review and incorporate this ever-growing wealth of clinical information.

Globalization of clinical research

Cardiovascular disease accounts for at least 30% of deaths worldwide (16.6 million people estimated in 2002). Notably, the majority of these individuals are in the low- and middle-income brackets, reaffirming that this is not only a disease of the rich. Projections by Beaglehole and Bonita indicate a growing cardiovascular burden across all income groups such that it is estimated that over three quarters of all deaths will occur secondary to chronic non-communicable disease by the year 2030. Of these non-communicable diseases, cardiovascular disease will be the most dominant. The socioeconomic determinants of this trend provide a compelling impetus to invest in research on health policy and integration of health systems that will enhance the application of available knowledge and close the treatment gaps that exist.

Remarkably, the cost of care bears little relationship to life expectancy: in this regard, the average per capita expenditure across a wide spectrum of countries is $2986 with an average life expectancy of approximately 79 years. At the extremes, Mexico spends slightly more than $800 and the United States in excess of $7000 per capita, yet both have below-average life expectancies indicating the complexity of this relationship. Notably, the Scandinavian countries, Canada, Switzerland, Australia, and France expend more than the median amounts but also have life expectancies in excess of the average.

The Treatment and Outcomes of Acute Coronary Syndromes in India (CREATE) registry offers insight into some of the challenges facing global cardiovascular research. It highlights the relatively young age at which MI occurs, the still dominant incidence of STEMI versus non-STEMI, with mortality from STEMI in excess of 8%. Remarkably, there is a delay from symptom onset to hospitalization of approximately 5 h for STEMI patients and an additional delay from hospitalization to fibrinolysis of nearly 1 h. The large majority of patients are transported to hospital by taxi or private vehicle, but as many as a third use public transportation and only a minority have access to ambulance transportation. As communicated by Prabhakaran, several factors impair research progress in India, including an entrenched bureaucracy, a lack of interdisciplinary and transdisciplinary research, resistance to change across all levels, substantial mobility and instability of the trained workforce, and the dominance of commercial contract research organizations (CROs) with a profit mandate.

On a broader global scale, perverse economic incentives exist in the provision of health care, and there remain huge disparities in access to high-quality health care. Moreover, the chasm between what we know versus how we integrate knowledge, coupled with fear of liability and a sometimes unreasonable quest for diagnostic certainty, contribute to inefficiencies. The treatment-risk paradox is pervasive, and too many dollars are spent on marginal gains or the so-called "flat portion" of the cost–benefit curve.

It is reassuring that there appears to be a renewed understanding of the importance of global academic collaboration based on several factors, including information technology and its transformation of the world into a global village. Moreover, there is a commonality of health-related issues and increasing concern about the costs of health care, which are driving an effort to acquire the best metrics for demonstration of return on investment in health care costs. An increasing number of questions regarding comparative efficacy that require head-to-head evaluations ensures no lack of meaningful projects to undertake. As mortality declines and life expectancy increases in a number of countries, new emphasis on better metrics to assess quality of life has emerged. Striking a balance between the content of care and elements associated with human behavior that contribute to the epidemics of obesity and diabetes remains a major challenge. In this regard, better understanding of the future of personalized medicine and genomics versus broad population approaches is mandatory. An important caveat for research in the developing world relates to statements by both the World Health Organization and the World Medical Organization affirming that, when conducting research in developing countries, it is necessary to ensure that the results of the research will be applicable to those populations in whom it is conducted so that they can benefit from the results.

As one surveys the global treatment gap, it is sobering to contemplate that <10% of global health research is devoted to diseases comprising 90% of the global disease burden. Indeed, a third of the world's population receives only 2% of global health resources, and only 5% of health research is devoted to prevention resources versus 95% dedicated to treatment. Daar et al. highlight six key challenges in tackling chronic non-communicable disease: (1) raising public awareness; (2) enhancing economic, legal, and environmental policies; (3) modifying risk factors; (4) engaging businesses and community; (5) mitigating health impacts of poverty and urbanization; and (6) reorienting health systems.

The VIGOUR Group is well prepared to execute its mission of enhancing worldwide cardiovascular health by the creation, implementation, and evaluation of novel strategies developed through global collaboration. Shared perspectives among group members relate to a strong social conscience and recognition of partnership within a global village. Not only is there an appreciation of the profound unmet needs that exist but also of the mismatch between resources on health expenditures versus key unanswered research questions. In a recent publication by Califf et al. from the VIGOUR Group, four key issues were identified: (1) the lack of definitive evidence to guide care, (2) disease heterogeneity, (3) inadequate funding, and (4) paucity of new leadership. To foster global academic collaboration, infrastructure at all health professional levels is needed, fiscal transparency and stability of academic research organizations (AROs) are required, and the right balance must be struck between individual versus group rewards for achievement. There is a compelling need to develop new leaders and define an appropriate career path for those engaging in these efforts. Each of these issues is associated with opportunities and strategies that will help to drive the cycle of quality on a global basis.

Role of AROs

Conducting high-quality global clinical research is increasingly challenging. The world is "flattening" (Thomas L. Freidman) due to a variety of forces including advances in information technology that allow efficient sharing of data across the globe. However, multiple impediments remain for efficient clinical trial conduct.

Most AROs have three key priorities: (1) patient care, (2) education, and (3) research. These priorities are often reflected in mission statements such as those from the Duke Clinical Research Institute (DCRI), the Brazilian Clinical Research Institute (BCRI), the Canadian VIGOUR Centre (CVC), and the Uppsala Clinical Research Center (UCR). The typical ARO will encompass a variety of research initiatives, including clinical trials, registries, health economics, quality-of-life projects, methodological research, core laboratories, and education. These programs are supported by a framework of coordinating center services. By contrast, missions of commercial CROs are different and typically reflect a goal of maximizing returns or providing efficient services. In a simplistic view, an ARO performs research, and a commercial CRO performs research services.

The United States and many parts of the world are experiencing a shortage of clinical trial investigators and coordinators. Financial pressures and the demands of clinical practice both are central issues of concern, along with growing complexities involving contracts and regulations, lack of training, and less infrastructure to support site-based research at many institutions. Efforts are needed to better understand local site challenges and to respond to those challenges. The Clinical Trials Network, which is part of the National Institutes of Health Roadmap (www.ctnbestpractices.org), provides site investigators with opportunities to learn and network in support of their daily activities.

Cardiovascular disease is and likely will remain the number one cause of death in the world; thus, identifying new and promising therapies is critical. Large clinical outcomes trials will remain the standard for assessing the benefits and safety of new agents, and, as such, clinical trials to evaluate novel therapies will remain large in size and require a global effort. Global clinical research driven by collaboration will be essential to complete these large trials quickly and efficiently. Relationships such as those that have been established between the DCRI and BCRI in Sao Paulo, Brazil, will help create the foundation and infrastructure for performing quality clinical research in the future. Several key priorities for AROs include: (1) creating a culture of excellence and partnership; (2) evaluating novel, efficient, and less costly trial designs and operations; (3) promoting evidence-based medicine and evidence-based trial operations; and (4) developing a sustainable clinical research community through focused support of site investigators. The future has challenges but also exciting opportunities.

Recommended reading

1. Ageno W, Garcia D et al (2009) Prevention and treatment of bleeding complications in patients receiving vitamin K antagonists, part 2: treatment. Am J Hematol 84(9):584–588
2. Anderson FA Jr, Zayaruzny M et al (2007) Estimated annual numbers of US acute-care hospital patients at risk for venous thromboembolism. Am J Hematol 82(9):777–782
3. Anderson J (2007) ACC/AHA 2007 guidelines for the management of patients with unstable angina/non-ST elevation myocardial infarction: executive summary. Circulation 116(7):803–877

4. Anderson WG, Alexander SC et al (2008) What concerns me is…' Expression of emotion by advanced cancer patients during outpatient visits. Support Care Cancer 16(7):803–811
5. Angiolillo DJ, Gibson CM et al (2011) Differential effects of omeprazole and pantoprazole on the pharmacodynamics and pharmacokinetics of clopidogrel in healthy subjects: randomized, placebo-controlled, crossover comparison studies. Clin Pharmacol Ther 89(1):65–74
6. Angiolillo DJ, Shoemaker SB et al (2007) Randomized comparison of a high clopidogrel maintenance dose in patients with diabetes mellitus and coronary artery disease: results of the Optimizing Antiplatelet Therapy in Diabetes Mellitus (OPTIMUS) study. Circulation 115(6):708–716
7. Ay C, Dunkler D et al (2010) Prediction of venous thromboembolism in cancer patients. Blood 116(24):5377–5382
8. Bailar JC III, Mosteller F (1988) Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations. Ann Intern Med 108(2):266–273
9. Beaglehole R, Bonita R (2008) Global public health: a scorecard. Lancet 372(9654):1988–1996
10. Becker RC (2002) Antithrombotic therapy after myocardial infarction. N Engl J Med 347(13):1019–1022
11. Berger JS, Frye CB et al (2008) Impact of clopidogrel in patients with acute coronary syndromes requiring coronary artery bypass surgery: a multicenter analysis. J Am Coll Cardiol 52(21): 1693–1701
12. Bhatt DL, Cryer BL et al (2010) Clopidogrel with or without omeprazole in coronary artery disease. N Engl J Med 363(20): 1909–1917
13. Blom JW, Doggen CJ et al (2005) Malignancies, prothrombotic mutations, and the risk of venous thrombosis. JAMA 293(6): 715–722
14. Brandt JT, Close SL et al (2007) Common polymorphisms of CYP2C19 and CYP2C9 affect the pharmacokinetic and pharmacodynamic response to clopidogrel but not prasugrel. J Thromb Haemost 5(12):2429–2436
15. Brazilian Clinical Research Institute (2009) Brazilian Clinical Research Institute. http://www.bcri.org.br/v2/. Accessed 8 Mar 2011
16. Breet NJ, van Werkum JW et al (2010) Comparison of platelet function tests in predicting clinical outcome in patients undergoing coronary stent implantation. JAMA 303(8):754–762
17. Califf RM (2008) The benefits of moving quality to a national level. Am Heart J 156(6):1019–1022
18. Califf RM, Armstrong PW et al (2010) Towards a new order in cardiovascular medicine: re-engineering through global collaboration. Eur Heart J 31(8):911–917
19. Camm AJ, Kirchhof P et al (2010) Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology. Europace 12(10):1360–1420
20. Chan MY, Cohen MG et al (2008) Phase 1b randomized study of antidote-controlled modulation of factor IXa activity in patients with stable coronary artery disease. Circulation 117(22):2865–2874
21. Chan MY, Weitz JI et al (2009) Catheter thrombosis and percutaneous coronary intervention: fundamental perspectives on blood, artificial surfaces and antithrombotic drugs. J Thromb Thrombolysis 28(3):366–380
22. Cohen AT, Tapson VF et al (2008) Venous thromboembolism risk and prophylaxis in the acute hospital care setting (ENDORSE study): a multinational cross-sectional study. Lancet 371(9610): 387–394
23. Cohen M, Gensini GF et al (2003) The safety and efficacy of subcutaneous enoxaparin versus intravenous unfractionated heparin and tirofiban versus placebo in the treatment of acute ST-segment elevation myocardial infarction patients ineligible for reperfusion (TETAMI): a randomized trial. J Am Coll Cardiol 42(8):1348–1356
24. Connolly S, Flaker G et al (2011) Apixaban in patients with atrial fibrillation. N Engl J Med 364(9):806–817
25. Connolly SJ, Ezekowitz MD et al (2009) Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med 361(12):1139–1151
26. Crowther MA, Warkentin TE (2009) Managing bleeding in anticoagulated patients with a focus on novel therapeutic agents. J Thromb Haemost 7(Suppl 1):107–110
27. Duke Clinical Research Institute (2009) "About Us". http://www.dcri.org/about-us. Accessed 8 Mar 2011
28. Eikelboom JW, Weitz JI (2008) Selective factor Xa inhibition for thromboprophylaxis. Lancet 372(9632):6–8
29. Eriksson BI, Borris LC et al (2008) Rivaroxaban versus enoxaparin for thromboprophylaxis after hip arthroplasty. N Engl J Med 358(26):2765–2775
30. Eriksson BI, Dahl OE et al (2007) Oral dabigatran etexilate versus subcutaneous enoxaparin for the prevention of venous thromboembolism after total knee replacement: the RE-MODEL randomized trial. J Thromb Haemost 5(11):2178–2185
31. Ferraris VA, Ferraris SP et al (2007) Perioperative blood transfusion and blood conservation in cardiac surgery: the Society of Thoracic Surgeons and The Society of Cardiovascular Anesthesiologists clinical practice guideline. Ann Thorac Surg 83(5 Suppl):S27–S86
32. Fitchett D, Eikelboom J et al (2009) Dual antiplatelet therapy in patients requiring urgent coronary artery bypass grafting surgery: a position statement of the Canadian Cardiovascular Society. Can J Cardiol 25(12):683–689
33. Friedman LM, Furberg CD, DeMets DL (eds) (1996) Fundamentals of clinical trials. Mosby-Year Book, St. Louis
34. Garcia D, Libby E et al (2010) The new oral anticoagulants. Blood 115(1):15–20
35. Geerts WH, Bergqvist D et al (2008) Prevention of venous thromboembolism: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th edition). Chest 133(6 Suppl):381S–453S
36. Gharacholou SM, Alexander KP et al (2010) Implications and reasons for the lack of use of reperfusion therapy in patients with ST-segment elevation myocardial infarction: findings from the CRUSADE initiative. Am Heart J 159(5):757–763
37. Gilard M, Arnaud B et al (2008) Influence of omeprazole on the antiplatelet action of clopidogrel associated with aspirin: the randomized, double-blind OCLA (Omeprazole CLopidogrel Aspirin) study. J Am Coll Cardiol 51(3):256–260
38. Glickman SW, McHutchison JG et al (2009) Ethical and scientific implications of the globalization of clinical research. N Engl J Med 360(8):816–823
39. Glynn RJ, Danielson E et al (2009) A randomized trial of rosuvastatin in the prevention of venous thromboembolism. N Engl J Med 360(18):1851–1861
40. Gurbel PA, Becker RC et al (2007) Platelet function monitoring in patients with coronary artery disease. J Am Coll Cardiol 50(19): 1822–1834
41. Hansen ML, Sorensen R et al (2010) Risk of bleeding with single, dual, or triple therapy with warfarin, aspirin, and clopidogrel in patients with atrial fibrillation. Arch Intern Med 170(16):1433–1441
42. Harrington RA, Becker RC et al (2008) Antithrombotic therapy for non-ST-segment elevation acute coronary syndromes: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th edition). Chest 133(6 Suppl):670S–707S
43. Heit JA (2008) The epidemiology of venous thromboembolism in the community. Arterioscler Thromb Vasc Biol 28(3):370–372

44. Hirsh J, Guyatt G et al (2008) Executive summary: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th edition). Chest 133(6 Suppl):71S–109S
45. Hirsh J, Guyatt G et al (2008) Antithrombotic and thrombolytic therapy: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th edition). Chest 133(6 Suppl): 110S–112S
46. Hlatky MA, Greenland P et al (2009) Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. Circulation 119(17):2408–2416
47. Kakkar AK, Brenner B et al (2008) Extended duration rivaroxaban versus short-term enoxaparin for the prevention of venous thromboembolism after total hip arthroplasty: a double-blind, randomised controlled trial. Lancet 372(9632):31–39
48. Keegan SP, Patrick DM et al (2009) Prevention of perioperative venous thromboembolism. Int Anesthesiol Clin 47(4):55–64
49. Khorana AA, Kuderer NM et al (2008) Development and validation of a predictive model for chemotherapy-associated thrombosis. Blood 111(10):4902–4907
50. Kreatsoulas C, Anand SS (2010) The impact of social determinants on cardiovascular disease. Can J Cardiol 26(Suppl C):8C–13C
51. Kushner FG, Hand M et al (2009) 2009 Focused updates: ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction (updating the 2004 guideline and 2007 focused update) and ACC/AHA/SCAI guidelines on percutaneous coronary intervention (updating the 2005 guideline and 2007 focused update) a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 54(23):2205–2241
52. Lassen MR, Ageno W et al (2008) Rivaroxaban versus enoxaparin for thromboprophylaxis after total knee arthroplasty. N Engl J Med 358(26):2776–2786
53. Lassen MR, Raskob GE et al (2009) Apixaban or enoxaparin for thromboprophylaxis after knee replacement. N Engl J Med 361(6): 594–604
54. Leissinger CA, Blatt PM et al (2008) Role of prothrombin complex concentrates in reversing warfarin anticoagulation: a review of the literature. Am J Hematol 83(2):137–143
55. Lopes RD, Alexander KP (2009) Antiplatelet therapy in older adults with non-ST-segment elevation acute coronary syndrome: considering risks and benefits. Am J Cardiol 104(5 Suppl):16C–21C
56. Lopes RD, Becker RC et al (2009) Highlights from the international symposium of thrombosis and anticoagulation in internal medicine, October 23–25, 2008, Sao Paulo, Brazil. J Thromb Thrombolysis 28(1):106–116
57. Lopes RD, Elliott LE et al (2009) Antithrombotic therapy and outcomes of patients with atrial fibrillation following primary percutaneous coronary intervention: results from the APEX-AMI trial. Eur Heart J 30(16):2019–2028
58. Lopes RD, Piccini JP et al (2008) Antithrombotic therapy in atrial fibrillation: guidelines translated for the clinician. J Thromb Thrombolysis 26(3):167–174
59. Matchar DB, Jacobson A et al (2010) Effect of home testing of international normalized ratio on clinical events. N Engl J Med 363(17):1608–1620
60. Mehta SR, Boden WE et al (2008) Antithrombotic therapy with fondaparinux in relation to interventional management strategy in patients with ST- and non-ST-segment elevation acute coronary syndromes: an individual patient-level combined analysis of the Fifth and Sixth Organization to Assess Strategies in Ischemic Syndromes (OASIS 5 and 6) randomized trials. Circulation 118(20):2038–2046
61. Nagarakanti R, Ezekowitz MD et al (2011) Dabigatran versus warfarin in patients with atrial fibrillation: an analysis of patients undergoing cardioversion. Circulation 123(2):131–136
62. O'Connell C, Razavi P et al (2011) Unsuspected pulmonary emboli adversely impact survival in patients with cancer undergoing routine staging multi-row detector computed tomography scanning. J Thromb Haemost 9(2):305–311
63. Oldgren J, Alings M et al (2010) Dabigatran versus warfarin in atrial fibrillation patients with low, moderate, and high $CHADS_2$ score: a RE-LY subgroup analysis. J Am Coll Cardiol 55(10A):A1.E2
64. Palumbo A, Rajkumar SV et al (2008) Prevention of thalidomide- and lenalidomide-associated thrombosis in myeloma. Leukemia 22(2):414–423
65. Pieper KS, Tsiatis AA et al (2004) Differential treatment benefit of platelet glycoprotein IIb/IIIa inhibition with percutaneous coronary intervention versus medical therapy for acute coronary syndromes: exploration of methods. Circulation 109(5):641–646
66. Santos AT, Kalil RA et al (2006) A randomized, double-blind, and placebo-controlled study with tranexamic acid of bleeding and fibrinolytic activity after primary coronary artery bypass grafting. Braz J Med Biol Res 39(1):63–69
67. Sibbing D, Braun S et al (2009) Platelet reactivity after clopidogrel treatment assessed with point-of-care analysis and early drug-eluting stent thrombosis. J Am Coll Cardiol 53(10):849–856
68. Siller-Matula JM, Jilma B et al (2010) Effect of proton pump inhibitors on clinical outcome in patients treated with clopidogrel: a systematic review and meta-analysis. J Thromb Haemost 8(12):2624–2641
69. Singer DE, Albers GW et al (2008) Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest 133(6 Suppl):546S–592S
70. Steg PG, FitzGerald G et al (2009) Risk stratification in non-ST-segment elevation acute coronary syndromes: troponin alone is not enough. Am J Med 122(2):107–108
71. Thygesen K, Alpert JS et al (2007) Universal definition of myocardial infarction. J Am Coll Cardiol 50(22):2173–2195
72. Wallentin L (2009) P2Y(12) inhibitors: differences in properties and mechanisms of action and potential consequences for clinical use. Eur Heart J 30(16):1964–1977
73. Wallentin L, Becker RC et al (2009) Ticagrelor versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 361(11):1045–1057
74. Wallentin L, Yusuf S et al (2010) Efficacy and safety of dabigatran compared with warfarin at different levels of international normalised ratio control for stroke prevention in atrial fibrillation: an analysis of the RE-LY trial. Lancet 376(9745):975–983
75. Wann LS, Curtis AB et al (2011) 2011 ACCF/AHA/HRS focused update on the management of patients with atrial fibrillation (update on dabigatran). A report of the American College of Cardiology Foundation Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 57(11): 1330–1337
76. Wiviott SD, Braunwald E et al (2008) Greater clinical benefit of more intensive oral antiplatelet therapy with prasugrel in patients with diabetes mellitus in the trial to assess improvement in therapeutic outcomes by optimizing platelet inhibition with prasugrel-Thrombolysis in Myocardial Infarction 38. Circulation 118(16):1626–1636
77. Wiviott SD, Braunwald E et al (2007) Prasugrel versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 357(20):2001–2015

78. Yusuf S (2001) Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. N Engl J Med 345(7):494–502
79. Yusuf S, Bosch J et al (2008) Sensible guidelines for the conduct of large randomized trials. Clin Trials 5(1):38–39
80. Yusuf S, Mehta SR et al (2006) Effects of fondaparinux on mortality and reinfarction in patients with acute ST-segment elevation myocardial infarction: the OASIS-6 randomized trial. JAMA 295(13):1519–1530
81. Yusuf S, Mehta SR et al (2006) Comparison of fondaparinux and enoxaparin in acute coronary syndromes. N Engl J Med 354(14):1464–1476

Warfarin and acetaminophen interaction: a summary of the evidence and biologic plausibility

Renato D. Lopes,[1] John D. Horowitz,[2] David A. Garcia,[3] Mark A. Crowther,[4] and Elaine M. Hylek[5]

[1]Duke Clinical Research Institute, Duke University Medical Center, Durham, NC; [2]University of Adelaide, Queen Elizabeth Hospital, Adelaide, Australia; [3]Department of Medicine, University of New Mexico Health Sciences Center, Albuquerque, NM; [4]Department of Medicine, McMaster University, Hamilton, ON; and [5]Department of Medicine, Research Unit, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA

Case presentation

Ms TS is a 66-year-old woman who receives warfarin for prevention of systemic embolization in the setting of hypertension, diabetes, and atrial fibrillation. She had a transient ischemic attack about 4 years ago when she was receiving aspirin. Her INR control was excellent; however, over the past few months it has become erratic, and her average dose required to maintain an INR of 2.0 to 3.0 appears to have decreased. She has had back pain over this same period and has been taking acetaminophen at doses as large as 650 mg four times daily, with her dose varying based on her symptoms. You recall a potential interaction and wonder if (1) her acetaminophen use is contributing to her loss of INR control, and (2) does this interaction place her at increased risk of warfarin-related complications?

Warfarin has remained the most commonly prescribed vitamin K antagonist (VKA) since its introduction into clinical practice approximately 60 years ago. VKAs exert their effect by inhibiting the cyclic replenishment of reduced vitamin K, an obligate cofactor in the γ-carboxylation of the biologically inactive procoagulant factors II, VII, IX, and X, as well as the anticoagulant factors protein C, protein S, and protein Z.[1] The resultant anticoagulant effect is measured by the international normalized ratio (INR), which for most indications is targeted between 2.0 and 3.0. Thromboembolism, major hemorrhage, and death have all been strongly linked to the proportion of time spent in this therapeutic range.[2,3] Despite its efficacy in preventing and treating thromboembolic disease, warfarin has several limitations that challenge its effectiveness in clinical practice, including a narrow therapeutic index, variable dose-response, and importantly the potential for important interactions with numerous commonly used medications.[4]

Reports of an interaction between warfarin and acetaminophen first appeared in the literature in 1968.[5] Acetaminophen is part of the class of drugs known as "aniline analgesics"; it is the only such drug still in use today.[6] Acetaminophen is used worldwide as an analgesic and antipyretic. Because aspirin and other nonsteroidal anti-inflammatory drugs inhibit platelet function and can cause injury to the gastric mucosa, acetaminophen is the analgesic of choice for patients receiving oral anticoagulant therapy. Establishing the validity of this interaction is critically important as acetaminophen is currently the recommended first-line therapy for pain control in older adults, the group at highest risk of hemorrhage and concomitant use of VKAs.[7] The objectives of this focused review are to summarize the observational and randomized data investigating this interaction, to provide insights into possible biologic mechanisms, and to suggest clinical practice recommendations for patients receiving both VKAs and acetaminophen.

Observational data

Early case reports demonstrated a temporal increase in the INR among persons taking warfarin after acetaminophen exposure, suggesting an interaction.[8,9] Subsequent observational studies investigated the relationships between acetaminophen and INR and acetaminophen and hemorrhage among persons prescribed warfarin. In a case-control study, 93 consecutive patients with an INR greater than 6 were compared with 196 randomly selected control patients with an INR in the range of 1.7 to 3.3. Participants were interviewed within 24 hours of the INR measurement, and pertinent exposures within the previous 7 days were recorded. Acetaminophen was independently associated in a dose-dependent fashion with an INR greater than 6.0. For persons taking acetaminophen 9.1 g/week or more, the odds of an INR greater than 6.0 were increased 10-fold. This dose-response relationship persisted after controlling for other factors known to potentiate warfarin.[10]

Another case-control study enrolled 53 patients with an INR greater than 4.5 and 106 control patients with an in-range INR. Amiodarone (9.4% vs 0%, $P < .004$), acetaminophen (18.9% vs 0.9%, $P < .001$), tramadol (5.6% vs 0%, $P < .04$), ofloxacine (11.3% vs 1.9%, $P < .001$), and lactulose (11.3% vs 0%, $P < .001$) were associated with INR elevation. Other factors included fever, malnutrition, dehydration, and acute diarrhea.[11] Bleeding complications occurred in 19.2% of cases versus 3.9% of the controls.

Shalansky et al prospectively studied 171 warfarin-treated patients to assess the risk of bleeding and elevated INR associated with the use of complementary and alternative medicines.[12] Patients kept a diary of selected exposures for 16 weeks. Pharmacy, laboratory, and medical records were subsequently queried for evidence of bleeding or elevations in INR. Acetaminophen was associated with increased risk of bleeding (OR = 1.42; 95% confidence interval [CI], 1.05-1.90) and INR elevation, although the latter did not achieve statistical significance (OR = 1.76; 95% CI, 0.85-3.63, $P = .13$).[12] Warfarin use of less than 3 months' duration was the only statistically significant risk factor identified for increased INR.

In a retrospective study of a postmortem toxicology database, Launiainen et al[13] reported an association of combination therapy with fatal hemorrhage. Of the 328 patients who were taking warfarin at the time of death, a potentially interacting drug was present in one-third, and acetaminophen was the most common (50%). Concomitant use of acetaminophen and warfarin was associated with a 4.6 and 2.7 times higher risk of fatal bleeding than either acetaminophen or warfarin alone, respectively.[13]

Submitted August 3, 2011; accepted August 30, 2011. Prepublished online as *Blood* First Edition paper, September 12, 2011; DOI 10.1182/blood-2011-08-335612.

Table 1. Randomized evaluations of the impact of acetaminophen on the pharmacodynamics of warfarin

Study, period	Study size	Acetaminophen dose	Study outcome
Antlitz,[16] 1968	20	650 mg vs placebo for 2 doses	No significant difference in prothrombin times between acetaminophen and placebo during the 48-hour period after 2 doses of study drug.
Mahé,[17] 2004	11	4 g/d vs placebo for 14 d	The mean observed INR was significantly increased after 4 days in the acetaminophen group. The mean maximum INR observed was 3.47 in the acetaminophen arm and 2.61 in the placebo arm ($P = .001$). The mean maximum increase in the INR was 1.04 (acetaminophen group) versus 0.20 (placebo group; $P = .003$).
Mahé,[18] 2006	20	4 g/d vs placebo for 14 d	The mean maximum INR observed was 3.45 in the acetaminophen arm and 2.66 in the placebo arm ($P = .03$). The mean maximum increase in the INR was 1.20 (acetaminophen group) versus 0.37 (placebo group; $P < .001$). Significant reductions in the vitamin K-dependent clotting factors II, VII, IX, and X accompanied the increase in INR.
Parra,[19] 2007	36	2 g/d, 4 g/d; vs placebo for 4 wk	At week 2, the 2 g/d group had significantly higher INR compared with placebo ($P = .01$). At weeks 1, 2, and 3, the 4 g/d group had significantly higher INR than placebo ($P = .04$, $P = .01$, and $P = .01$, respectively).
Zhang,[20] 2011	45	2 g/d, 3 g/d; vs placebo for 10 d	The mean INR increase was 0.70, 0.67, and 0.14 in the 2 g/d group, 3 g/d group, and in the placebo group, respectively ($P = .01$). Factor VII levels were lower in the acetaminophen groups compared with placebo, but not levels of factors II and V.

INR indicates international normalized ratio.

No association between acetaminophen and INR was found in a retrospective study of 54 persons prescribed nonwarfarin VKAs, acenocoumarol, or phenprocoumon. However, 9 patients in the acetaminophen group required a reduction in coumarin dose after acetaminophen exposure compared with one in the control group, which may have blunted any increase in the INR.[14]

Observational studies investigating the interaction between acetaminophen and warfarin are limited in their ability to ascertain duration, dose, and cumulative dose of acetaminophen given the nonprescription or "over-the-counter" status of acetaminophen. The temporal relationship between acetaminophen exposure and INR elevation mandates standardized INR measurement at baseline and at prespecified intervals among persons already stabilized on warfarin. These limitations and the concern for residual confounding mandate a randomized assessment.

Intervention and randomized data

To investigate an observed increase in INR after a course of acetaminophen, a 74-year-old man was rechallenged with acetaminophen 1 g 4 times per day for 3 consecutive days. On rechallenge, the INR increased from 2.4 to 6.3, factor VII activity decreased from 29.4% to 15.5%, and there was no significant change in warfarin plasma concentration (1.54 μg/mL vs 1.34 μg/mL). These findings argued against a pharmacokinetic basis for the interaction and, instead, suggested a pharmacodynamic mechanism for the elevation in INR.[15]

Five randomized trials have been performed to evaluate the effect of acetaminophen on INR in patients treated with warfarin (Table 1). Four of the 5 studies were positive. The one negative study, published in 1969, randomized 20 patients with stable prothrombin times to 2 doses of acetaminophen 650 mg or matching placebo given at 8 am and 12 noon. Prothrombin times were measured at 8 am, 10 am, 12 noon, 2 pm, and 4 pm on the day of study drug administration, and at 8 am on the following 2 days. This study found no significant difference in prothrombin times between acetaminophen and placebo during the 48-hour period after study drug administration. This study demonstrated either no effect or no measurable effect of low-dose acetaminophen given for a very short duration.[16]

In 2004, Mahé et al randomized 11 stable warfarin patients to 4 grams per day of acetaminophen or placebo for 14 days.[18] This was a prospective, double-blind, placebo-controlled trial. The mean observed INR was significantly increased after 4 days and throughout the study period in the acetaminophen group, whereas no differences were observed in the placebo group ($P = .001$, Table 1).[17] In a follow-up study 2 years later, the authors conducted a double-blind, placebo-controlled, randomized, crossover study in 20 patients with stable INR. Participants were randomized to receive placebo or acetaminophen 1 g 4 times daily for 14 days. INR and clotting factor activities were measured before the first drug administration and then on days 2, 4, 7, 9, 11, and 14. The authors demonstrated that the mean INR rose rapidly after the start of acetaminophen and was significantly increased after 1 week of acetaminophen intake compared with placebo ($P = .0002$). Significant reductions in the vitamin K-dependent clotting factors II, VII, IX, and X accompanied this increase in INR.[18]

In 2007, Parra et al performed another randomized, double-blind, placebo-controlled trial testing the effect of different doses of acetaminophen versus placebo on INR in patients stabilized on warfarin.[19] Patients received acetaminophen 2 g/day (n = 12) or 4 g/day (n = 12) or matching placebo 4 times/day (n = 12) for 4 weeks. More than 50% of the patients receiving acetaminophen exceeded the upper limit of the therapeutic INR range compared with 17% in the placebo group. At week 4, no differences were observed in alanine aminotransferase and aspartate aminotransferase between either of the acetaminophen groups or placebo. Patients receiving 4 g/day had significantly higher alanine aminotransferase at week 2 compared with placebo, suggesting a modest and temporary hepatic effect associated with the higher dose of acetaminophen. In more than 80% of the patients who developed an elevated INR and did not have their dose adjusted, the INR returned to normal when acetaminophen was stopped.[19]

Finally and more recently, the largest trial to date randomized 45 patients to one of 3 arms: acetaminophen 2 g/day, acetaminophen 3 g/day, and placebo for 10 consecutive days. Both doses of acetaminophen were associated with an increase in INR compared with placebo. The maximum INR increase was independently associated with a decrease in factor II ($P < .001$) and factor VII ($P < .001$) activities with an increase in acetaminophen plasma concentrations ($P < .001$).[20]

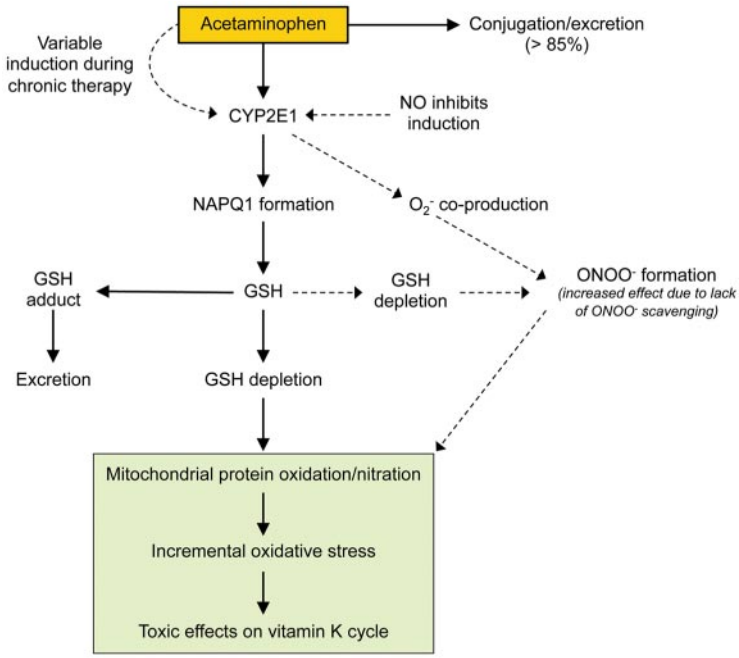


Figure 1. Determinants of variable induction of oxidative stress by acetaminophen. GSH indicates glutathione; NO, nitric oxide; and $ONOO^-$, peroxynitrite.

Biologic plausibility and mechanistic insights

A number of investigators have performed carefully designed experiments to evaluate the possibility that acetaminophen and/or its metabolites affect warfarin pharmacokinetics. Given that warfarin is a racemic mixture of R- and S-enantiomers with substantially differing pharmacokinetics and potency, these investigations have examined both enantiomers and have been adequately powered to exclude any major interactions. Given the lack of evidence for a pharmacokinetic interaction, a pharmacodynamic mechanism was hypothesized.[21] The concept that acetaminophen might interact with warfarin by potentiating its inhibition of components of the vitamin K cycle was initially raised by Thijssen et al.[22] These investigators drew attention to the recent finding that acetaminophen overdose had been associated with elevation of INR and diminution of vitamin K-dependent factors VII and IX levels (and not factor VIIIc) in the absence of other indices of acetaminophen-induced hepatotoxicity as evidence that acetaminophen and/or its metabolites might inhibit vitamin K function.[23]

These investigators therefore evaluated the effects of acetaminophen and its metabolite N-acetyl-p-benzoquinone-imine (NAPQ1) on the activity of 2 key enzymes of the vitamin K cycle, vitamin K-dependent carboxylase and vitamin K-epoxide reductase (VKOR), in washed microsomal preparations. Acetaminophen did not affect activities of either enzyme. However, NAPQI oxidized vitamin K-hydroquinone (KH2), the "active" form of the vitamin. In addition, NAPQI directly inhibited vitamin K-dependent carboxylation. Furthermore, VKOR activity was inhibited by NAPQI. Therefore, NAPQI disrupted the vitamin K cycle, potentially at 3 sites. Although the potency of NAPQI as an inhibitor at any one point of the cycle appeared limited, the interactions are important as they may be synergistic and dependent on localized intracellular increases in concentrations of NAPQI under certain circumstances.

Implications of variable production of NAPQI with therapeutic acetaminophen ingestion

The production of NAPQI as a toxic metabolite of acetaminophen has received considerable scientific attention in the context of acetaminophen overdose. NAPQI production reflects largely or entirely metabolism of acetaminophen by cytochrome P4502E1 (CYP2E1) with substantially increased generation of CYP2E1 occurring during overdose.[24-26] Furthermore, there is considerable evidence that NAPQI depletes tissue sulfhydryls, including glutathione, and is covalently protein-bound.[27]

A number of metabolic pathways for acetaminophen have been delineated, including conjugation with glucuronic acid and subsequent elimination of the nontoxic conjugate. NAPQI generation, catalyzed by CYP2E1 in the presence of nicotinamide adenine dinucleotide phosphate (NADPH), does not inevitably lead to its accumulation, given that NAPQI can be rapidly cleared by conjugation with glutathione. The potential for NAPQI accumulation and toxicity therefore results from induction of CYP2E1 and/or depletion of glutathione. It is important to emphasize that variable generation of NAPQI reflects not only tissue acetaminophen concentrate but also induction of CYP2E1. Acetaminophen itself increases expression of CYP2E1, and this may occur with subtoxic doses of acetaminophen.[28,29] Other factors that may potentially induce CYP2E1 include ethanol and diabetes mellitus/hyperglycemia.[30,31] Other sources of variability in CYP2E1 activity include the Dra I polymorphism of the CYP2E1 gene, which may potentiate activity of CYP2E1 in response to inducing agents, and the inhibitory effect of nitric oxide (Figure 1).[31,32]

The effects of NAPQI in inactivating vitamin K-dependent γ-carboxylase and VKOR are not its only enzymatic interactions. NAPQI may also inhibit components of the mitochondrial electron transport chain.[33] However, detailed evaluation of the full extent of direct toxic effects of NAPQI has been limited to date.

Potential downstream effectors of CYP2E1 activation/NAPQI production by acetaminophen

The potential ramifications of CYP2E1 activation and NAPQI production on the vitamin K cycle are extensive and include: (1) inactivation of VKOR via oxidation of essential cysteine moieties, (2) impairment of reductive reactivation of VKOR, and (3) impairment of VKOR-supported activation of vitamin K γ-carboxylase. It is probable that, although the extent of CYP2E1 activation is critical to impairment of the vitamin K cycle in the presence of acetaminophen, NAPQI is not the

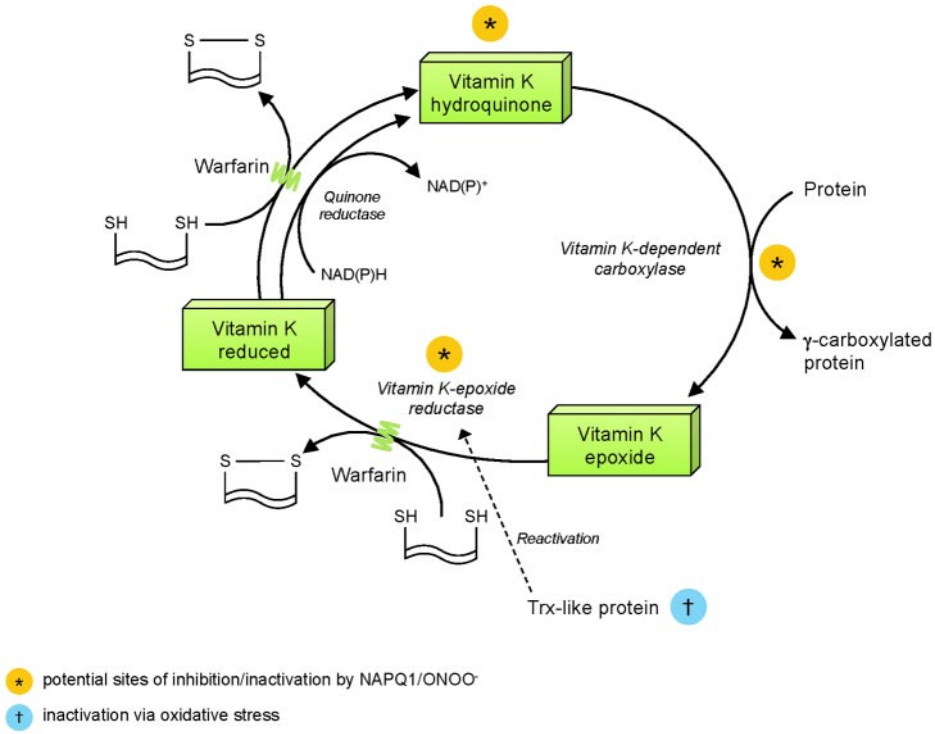


Figure 2. Points of potential disruption of vitamin K cycle and potentiation of warfarin effect by NAPQ1 and ONOO⁻. NAD(P)H indicates reduced nicotinamide adenine dinucleotide phosphate; NAD(P)⁺, oxidized nicotinamide adenine dinucleotide phosphate; ONOO⁻, peroxynitrite; and SH, sulfhydryl group.

only effector of the extensive oxidative changes that underlie this impairment. One such potential effector is peroxynitrite, a reactive species produced via the reaction of superoxide anion with nitric oxide. There is an extensive literature suggesting that peroxynitrite modulates the development of acetaminophen hepatotoxicity.[34] Release of reactive oxygen species via CYP2E1 activation has been shown to deplete sulfhydryl sources, such as reduced glutathione in mitochondria and endoplasmic reticulum.[30] Similarly, there is some evidence that scavenging of peroxynitrite may limit the cytotoxic effects of acetaminophen.[35]

The impact of CYP2E1 up-regulation is also potentially modified in other ways. The role of nitric oxide is particularly complex, as nitric oxide is both involved in peroxynitrite formation as well as, apparently, via its activation of soluble guanylate cyclase, able to limit CYP2E1-related toxicity by limiting its expression.[31,36] There is also considerable evidence that the transcription factor Nrf2, which controls antioxidant defense in part via increased glutathione synthesis, limits CYP2E1 toxicity.[37]

The activity of VKOR, vital to the integrity of the vitamin K cycle, is physiologically inhibited by oxidation of key cysteine moieties.[38,39] Thus, any form of oxidative stress, via NAPQI, peroxynitrite, or both, could inactivate VKOR, as originally demonstrated by Thijssen et al.[22] Indeed, depletion of glutathione and other sulfhydryl molecules is a common modality of peroxynitrite toxicity. Equally important is the susceptibility of VKOR reactivation to oxidative stress. Although the molecules reactivating VKOR have not been identified conclusively, there is considerable evidence that they are thioredoxin-like.[39] As activity of thioredoxin is itself impaired in the presence of oxidative stress and/or via its physiologic antagonist thioredoxin-interacting protein, it seems likely, although as yet unexplored, that this represents a further site of the acetaminophen-vitamin K interaction. It has also recently been demonstrated that thioredoxin activity drives the role of VKOR in supporting vitamin K-dependent γ-carboxylation.[39] Furthermore, the activity of vitamin K-dependent carboxylase is inhibited by oxidation of sulfhydryl groups, although these lie outside the catalytic site of the molecule.[40] Therefore, the observations of Thijssen et al[22] may reflect interplay of CYP2E1 activation, NAPQI and peroxynitrite production, and thioredoxin inactivation, as outlined in Figure 2.

General recommendations

The requisite features of causality exist for a warfarin/acetaminophen interaction: temporal relationship, measurable effect with dechallenge and rechallenge, dose-response, exclusion or accounting of other possible etiologic factors, and biologic plausibility. The strength of our clinical practice recommendations is low because, although the evidence of an important warfarin/acetaminophen interaction that results in INR variation is strong, there are no prospective management studies to indicate that the recommendations we make would reduce patient-important events, such as major bleeding or thrombosis.

In warfarin-treated patients who will use more than or equal to 2 g/day of acetaminophen for at least 3 consecutive days, we suggest that the INR should be tested 3 to 5 days after the first acetaminophen dose (grade 2C). In warfarin-treated patients with otherwise unexplained INR variability, acetaminophen use should be considered as a possible contributing factor (grade 2C).

On a wider scale, acetaminophen may disrupt, not only the production of the vitamin K-dependent proteins of the coagulation cascade, but all vitamin K-dependent proteins, such as those that normally function as inhibitors of calcification and modulate signal transduction and cell growth. Evaluation of long-term effects of acetaminophen ingestion with these changes in mind seems appropriate.

Acknowledgments

This work was supported by the Duke Clinical Research Institute, Durham, NC.

Authorship

Contribution: R.D.L. conducted the literature search; R.D.L. and J.D.H. wrote the first draft of the manuscript; and all authors provided clinical input and critical review of the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Renato D. Lopes, Duke Clinical Research Institute, Box 3850, 2400 Pratt St, Rm 0311, Terrace Level, Durham, NC 27705; e-mail: renato.lopes@duke.edu.

References

1. Ansell J, Hirsh J, Hylek EM, Jacobson A, Crowther M, Palareti G. Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th ed). *Chest.* 2008;133(6 suppl):160S-198S.
2. Connolly SJ, Pogue J, Eikelboom J, et al. ACTIVE W Investigators: benefit of oral anticoagulant over antiplatelet therapy in atrial fibrillation depends on the quality of international normalized ratio control achieved by centers and countries as measured by time in therapeutic range. *Circulation.* 2008;118(20):2029-2037.
3. White HD, Gruber M, Feyzi J, et al. Comparison of outcomes among patients randomized to warfarin therapy according to anticoagulant control: results from SPORTIF III and V. *Arch Intern Med.* 2007;167(3):239-245.
4. Wittkowsky AK, Boccuzzi SJ, Wogen J, Wygant G, Patel P, Hauch O. Frequency of concurrent use of warfarin with potentially interacting drugs. *Pharmacotherapy.* 2004;24(12):1668-1674.
5. Antlitz AM, Mead JA, Tolentino MA. Potentiation of oral anticoagulant therapy by acetaminophen. *Curr Ther Res.* 1968;10(10):501-507.
6. Bertolini A, Ferrari A, Ottani A, Guerzoni S, Tacchi R, Leone S. Paracetamol: new vistas of an old drug. *CNS Drug Rev.* 2006(Fall-Winter);12(3):250-275.
7. American Geriatrics Society Panel on Pharmacological Management of Persistent Pain in Older Persons. Pharmacological management of persistent pain in older persons. *J Am Geriatr Soc.* 2009;57(8):1331-1346.
8. Orme M, Breckenridge A, Cook P. Warfarin and distalgesic interaction. *Br Med J.* 1976;1(6003):200.
9. Bartle WR, Blakely JA. Potentiation of warfarin anticoagulation by acetaminophen. *JAMA.* 1991;265(10):1260.
10. Hylek EM, Heiman H, Skates SJ, Sheehan MA, Singer DE. Acetaminophen and other risk factors for excessive warfarin anticoagulation. *JAMA.* 1998;279(9):657-662.
11. Chatap G, Giraud K, Delay V, et al. Factors of therapeutic imbalance of antivitamins K and their haemorrhagic consequences in elderly. *Therapie.* 2008;63(5):393-401.
12. Shalansky S, Lynd L, Richardson K, Ingaszewski A, Kerr C. Risk of warfarin-related bleeding events and supratherapeutic international normalized ratios associated with complementary and alternative medicine: a longitudinal analysis. *Pharmacotherapy.* 2007;27(9):1237-1247.
13. Launiainen T, Sajantila A, Rasanen I, Vuori E, Ojanperä I. Adverse interaction of warfarin and paracetamol: evidence from a post-mortem study. *Eur J Clin Pharmacol.* 2010;66(1):97-103.
14. van den Bemt PM, Geven LM, Kuitert NA, Risselada A, Brouwers JR. The potential interaction between oral anticoagulants and acetaminophen in everyday practice. *Pharm World Sci.* 2002;24(5):201-204.
15. Gebauer MG, Nyfort-Hansen K, Henschke PJ, Gallus AS. Warfarin and acetaminophen interaction. *Pharmacotherapy.* 2003;23(1):109-112.
16. Antlitz AM, Awalt LF. A double-blind study of acetaminophen used in conjunction with oral anticoagulant therapy. *Curr Ther Res Clin Exp.* 1969;11(6):360-361.
17. Mahé I, Bertrand N, Drouet L, et al. Paracetamol: a haemorrhagic risk factor in patients on warfarin. *Br J Clin Pharmacol.* 2005;59(3):371-374.
18. Mahé I, Bertrand N, Drouet L, et al. Interaction between paracetamol and warfarin in patients: a double-blind, placebo-controlled, randomized study. *Haematologica.* 2006;91(12):1621-1627.
19. Parra D, Beckey NP, Stevens GR. The effect of acetaminophen on the international normalized ratio in patients stabilized on warfarin therapy. *Pharmacotherapy.* 2007;27(5):675-683.
20. Zhang Q, Bal-dit-Sollier C, Drouet L, et al. Interaction between acetaminophen and warfarin in adults receiving long-term oral anticoagulants: a randomized controlled trial. *Eur J Clin Pharmacol.* 2011;67(3):309-314.
21. Kwan D, Bartie WR, Walker SE. The effects of acetaminophen on pharmacokinetics and pharmacodynamics of warfarin. *J Clin Pharmacol.* 1999;39(1):68-75.
22. Thijssen HH, Soute BA, Vervoort LM, Claessens JG. Paracetamol (acetaminophen) warfarin interaction: NAPQI, the toxic metabolite of paracetamol, is an inhibitor of enzymes in the vitamin K cycle. *Thromb Haemost.* 2004;92(4):797-802.
23. Whyte IM, Buckley NA, Reith DM, Goodhew I, Seldon M, Dawson AH. Acetaminophen causes an increased international normalized ratio by reducing functional factor VII. *Ther Drug Monit.* 2000;22(6):742-748.
24. Patten CJ, Thomas PE, Guy RL, et al. Cytochrome P450 enzymes involved in acetaminophen activation by rat and human liver microsomes and their kinetics. *Chem Res Toxicol.* 1993;6(4):511-518.
25. Raucy JL, Lasker JM, Lieber CS, Black M. Acetaminophen activation by human liver cytochromes P450IIE1 and P450IA2. *Arch Biochem Biophys.* 1989;271(2):270-283.
26. Hinson JA, Roberts DT, James LP. Mechanisms of acetaminophen-induced liver necrosis. *Handb Exp Pharmacol.* 2010;196369-405.
27. James LP, Maveaux PR, Hinson JA. Acetaminophen-induced hepatotoxicity. *Drug Metab Dispos.* 2003;31(12):1499-1506.
28. Das J, Ghosh J, Manna P, Sil PC. Acetaminophen induced acute liver failure via oxidative stress and JNK activation: protective role of taurine by the suppression of cytochrome P450 2E1. *Free Radic Res.* 2010;44(3):340-355.
29. Kim SN, Seo JY, Jung da W, Lee MY, Jung YS, Kim YC. Induction of hepatic CYP2E1 by a subtoxic dose of acetaminophen in rats: increase in dichloromethane metabolism and carboxyhemoglobin elevation. *Drug Metab Dispos.* 2007;35(10):1754-1758.
30. Knockaert L, Descatoire V, Vadrot N, Fromenty B, Robin MA. Mitochondrial CYP2E1 is sufficient to mediate oxidative stress and cytotoxicity induced by ethanol and acetaminophen. *Toxicol In Vitro.* 2011;25(2):475-484.
31. Schafer A, Galuppo P, Fraccarollo D, et al. Increased cytochrome P4502E1 expression and altered hydroxyeicosatetraenoic acid formation mediate diabetic vascular dysfunction: rescue by guanylyl-cyclase activation. *Diabetes.* 2010;59(8):2001-2009.
32. Wong NA, Rae F, Simpson KJ, Murray GD, Harrison DJ. Genetic polymorphisms of cytochrome p4502E1 and susceptibility to alcoholic liver disease and hepatocellular carcinoma in a white population: a study and literature review, including meta-analysis. *Mol Pathol.* 2000;53(2):88-93.
33. Donnelly PJ, Walker RM, Racz WJ. Inhibition of mitochondrial respiration in vivo is an early event in acetaminophen-induced hepatotoxicity. *Arch Toxicol.* 1994;68(2):110-118.
34. Hinson JA, Reid AB, McCullough SS, James LP. Acetaminophen-induced hepatotoxicity: role of metabolic activation, reactive oxygen species, and mitochondrial permeability transition. *Drug Metab Rev.* 2004;36(3):805-822.
35. Bait ML, Knight TR, Farhood A, Jaeschke H. Scavenging peroxynitrite with glutathione promotes regeneration and enhances survival during acetaminophen-induced liver injury in mice. *J Pharmacol Exp Ther.* 2003;307(1):67-73.
36. Wu D, Cederbaum A. Nitric oxide donors prevent while the nitric oxide synthase inhibitor L-NAME increases arachidonic acid plus CYP2E1-dependent toxicity. *Toxicol Appl Pharmacol.* 2006;216(2):282-292.
37. Cederbaum A. Nrf2 and antioxidant defense against CYP2E1 toxicity. *Expert Opin Drug Metab Toxicol.* 2009;5(10):1223-1244.
38. Schulman S, Wang B, Weikal L, Rapoport TA. Vitamin K epoxide reductase prefers ER membrane-anchored thioredoxin-like redox partners. *Proc Natl Acad Sci U S A.* 2010;107(34):15027-15032.
39. Rishavy MA, Usubalieva A, Hallgren KW, Berkner KL. Electron relay through Cys43 and Cys 51 reduces VKOR to allow vitamin K reduction and facilitation of vitamin K-dependent protein carboxylation. *J Biol Chem.* 2011;286(9):7267-7278.
40. Tie JK, Jin DY, Loiselle DR, Pope RM, Straight DL, Stafford DW. Chemical modification of cysteine residues is a misleading indication of their status as active site in the vitamin K-dependent gamma-glutamyl carboxylation reaction. *J Biol Chem.* 2004;279(52):54079-54087.

Unique cases written by pediatric residents with a faculty member.

Under Pressure for a Diagnosis

Clinical Pediatrics
50(2) 166–168

DOI: 10.1177/0009922809350220
http://clp.sagepub.com
SAGE

Karsten Lunze, MD, MPH,[1] and Leonard Feldman, MD[2]

Patient Report

A 16-year-old adolescent male presented to his primary care physician with a low-grade fever and symptoms consistent with an upper respiratory tract infection. Over the next 3 months, he received 2 courses of azithromycin and albuterol. His cough never resolved, and he developed worsening dyspnea on exertion. He was finally admitted to an outside hospital, where a chest X ray revealed bilateral pleural effusions and an enlarged heart. Pleural and pericardial effusions were confirmed on CT.

His dyspnea on exertion and radiological findings prompted the placement of bilateral chest tubes followed by mechanical ventilation for 5 days. After an echocardiogram, a pericardiocentesis without concurrent cardiac catheterization drained 127 mL of fluid. No active bacterial, fungal, or viral infections were found. He was also treated with stress-dose steroids. He was transferred to Johns Hopkins Hospital with massive pleural fluid output of 3 to 5 L per day.

His past medical history was significant for ADHD, gastroesophageal reflux disease, milk protein allergy, and allergic colitis as an infant. He had been diagnosed with constitutional growth delay with, his height and weight were below the fifth percentile. His family history and social history were unremarkable.

On physical exam, he appeared fatigued and had an increased work of breathing. His respiratory rate was 23 breaths per minute with a room air oxygen saturation of 100%. His heart rate was 125 bpm, with a normal blood pressure. The lower lung breath sounds were decreased bilaterally. Jugular venous distention was not visible. He had a quiet precordium with a normal heart exam but for tachycardia. His abdomen was distended and demonstrated shifting dullness. He had no lower-extremity or sacral edema.

Pleural fluid studies revealed a pleural fluid-to-serum LDH (Lactate dehydrogenase) ratio of less than 0.6, fulfilling Light's criteria characteristic of transudate.[1] LDH was 124 U/L, less than two thirds the upper limit of normal. The pleural fluid-to-serum protein level was just greater than 0.5 at the outside hospital but was less than 0.5 at Johns Hopkins. Laboratory studies at the outside hospital and at Johns Hopkins were remarkable for a negative infectious disease workup and no signs of rheumatological disease or oncological disease. His liver function tests were abnormal on admission (Table 1).

An ECG, an echocardiogram, a thoracic and abdominal CT, and an abdominal ultrasound revealed a low-voltage QRS, a small pericardial effusion, bilateral areas of consolidation in the left and right lower lobes, large amounts of ascites, and patent hepatic vasculature with normal liver echotexture. A trial of stress-dose steroids did not improve his symptoms as 3 to 5 L/d of pleural fluid continued to drain from his chest tubes daily.

A follow-up echocardiogram 1 week later revealed impaired diastolic dysfunction, which led to a right and left heart catheterization with pressures consistent with constrictive pericarditis. A retrospective review of the first echocardiogram showed an increased diastolic inflow velocity E' by tissue Doppler and a diastolic septal bounce consistent with diastolic dysfunction and constrictive pericarditis. A right and left heart catheterization showed elevated end-diastolic pressures.

The diagnosis was effusive-constrictive pericarditis

Hospital Course

An uncomplicated pericardiotomy was performed. The procedure revealed chronically inflamed, fibrotic, and thickened parietal and visceral pericardial layers bridged by synechiae. Once the thick peel of visceral pericardium was incised, the freed right ventricle immediately bulged. At the end of the pericardiectomy, only the parietal pericardium on the diaphragmatic surface and posterior to the phrenic nerves was left. The patient's cardiac index as well as the central venous, pulmonary artery, and wedge pressures improved (Table 2). Pathological testing ruled

[1]Johns Hopkins Children's Center, Baltimore, MD, USA
[2]Johns Hopkins Hospital, Baltimore, MD, USA

Corresponding Author:
Karsten Lunze, Harriet Lane Pediatric Residency Program, Johns Hopkins Children's Center, 600 North Wolfe Street, CMSC 2-124, Baltimore, MD 21287-3224, USA
Email: Karsten.Lunze@post.harvard.edu

Table 1. Results of Laboratory Studies That Ruled Out an Infectious Etiology for the Polyserositis

Laboratory Test	Result
C-reactive protein	0.1 mg/dL
Erythrocyte sedimentation rate	2 mm/h
ANA screen	Negative
ANCA	Negative
Anti-DNA antibody	Negative
Adenovirus PCR	Negative
Chlamydia PCR	Negative
Coxsackie virus antibody	Negative
ECHO virus antibody	Negative
Enterovirus PCR	Negative
HIV antibody	Negative
Legionella PCR	Negative
Lyme antibody	Negative
Mycoplasma pneumoniae PCR	Negative
AST	36 U/L
ALT	62 U/L
Total bilirubin	3.4 mg/dL

NOTES: PCR= polymerase chain reaction; ANA=Anti-nuclear antibodies; ANCA=Anti-neutrophilic cytoplasmic antibodies; AST= Aspartate Aminotransferase; ALT= Alanine Aminotransferase.

Table 2. Cardiac Indices and Pressures Before and Immediately After Pericardiectomy

	Preoperative	Postoperative
Central venous pressure in mm Hg	35	14
Pulmonary artery pressure in mm Hg	40	25
Wedge pressure in mm Hg	25	8
Cardiac index in L/min per m^2 body surface	1.5	4.0

out an infectious etiology of the pericarditis. Both chest tubes were removed successfully 12 days later. The patient returned to his premorbid state 3 months after surgery.

Commentary

We describe a 16-year-old adolescent who presented with extensive third-space effusions and dramatic chest tube output attributable to effusive-constrictive pericarditis. Constrictive pericarditis is characterized by the persistence of elevated right atrial pressures after removal of pericardial fluid.[2] It is rare in children[3] and adults[4] and can remain unsuspected until autopsy. In our patient, pericarditis of unknown etiology developed chronically. In the absence of neoplastic processes, most cases of pericarditis are idiopathic in adults.[5]

The inflammatory pericardial effusion and the constriction as a result of diseased pericardium led to an impaired myocardial relaxation and diastolic dysfunction. The poor filling capability resulted in a decreased cardiac output with compensatory tachycardia. As a consequence of this gradual process, heart failure led to ascites, massive pleural effusions, and congestive hepatopathy.

Although it is a rare entity, of effusive-constrictive pericarditis since a visceral pericardiectomy is indicated, which is much more complex than a parietal pericardiectomy. However, drainage of the pericardial fluid or removal of only the parietal pericardium is ineffective when a visceral pericardial constriction is present.[2]

The diagnostic accuracy of conventional flow Doppler is poor in the case of constrictive processes.[4] Tissue Doppler imaging (TDI) is a novel, noninvasive technique for the assessment of cardiac function by myocardial tissue velocity and deformation rather than blood flow.[6] In contrast to flow Doppler, it allows us to distinguish constrictive physiology caused by pericardial pathology from restrictive physiology of diseased myocardium. Constrictive physiology impairing ventricular relaxation results in higher TDI e' waves at the septal in comparison to the lateral mitral ring respiratory septal motion shift as well. Current guidelines recommend TDI studies to diagnose diastolic dysfunction.[6] In our case, the detection of diastolic dysfunction by TDI ultimately led to the correct diagnosis of constrictive pericarditis.

The diagnosis of this treatable disease will be delayed if one fails to consider that a constrictive process can present concurrently with a tense effusion. Effusive-constrictive pericarditis should be included in the differential diagnosis of breathlessness in constellation with pleural effusions, pericardial effusions, and/or ascites, and should be investigated by TDI. Right- and left-cardiac catheterization during pericardiocentesis is still the diagnostic gold standard. However, characteristic TDI wave patterns can identify constrictive physiology and could render purely diagnostic cardiac catheterization redundant.

After pericardiocentesis, intrapericardial pressures decrease in patients with effusive-constrictive pericarditis, but their right atrial and right- and left-ventricular end-diastolic pressures remain elevated, although slightly reduced. This is known as dip-plateau morphology.

Conclusion

The clinical manifestations of effusive-constrictive pericarditis may be extracardial. Transudative pleural effusions should prompt an evaluation for cardiac diastolic dysfunction among other entities. For the detection of diastolic function, TDI can be an invaluable, noninvasive diagnostic tool and can distinguish restrictive

patterns from constrictive ones. Finally, effusive-constrictive pericarditis is an elusive diagnosis if pericardiocentesis occurs without concurrent heart catheterization. Consider effusive-constrictive pericarditis in patients without symptom resolution after pericardiocentesis.

Acknowledgments

We thank the residents and attendings at the Johns Hopkins Children's Center. We thank Fatima Lunze, MD, PhD, for useful discussions on tissue Doppler imaging. This article is dedicated to my clinical teacher Dr Ingram Schulze-Neick.

Declaration of Conflicting Interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The authors received no financial support for the research and/or authorship of this article.

References

1. Light RW, Macgregor MI, Luchsinger PC, Ball WC Jr. Pleural effusions: the diagnostic separation of transudates and exudates. *Ann Intern Med*. 1972;77:507-513.
2. Sagrista-Sauleda J, Angel J, Sanchez A, Permanyer-Miralda G, Soler-Soler J. Effusive-constrictive pericarditis. *N Engl J Med*. 2004;350:469-475.
3. Fowler N, Bove KE, Dunbar S, Meyer R. Clinical pathologic conference: fatigue, dyspnea and abdominal swelling in a 13-year-old boy. *Am Heart J*. 1978;96:533-542.
4. Marshall A, Ring N, Lewis T. Constrictive pericarditis: lessons from the past five years' experience in the South West Cardiothoracic Centre. *Clin Med*. 2006;6:592-597.
5. Sagrista-Sauleda J, Merce J, Permanyer-Miralda G, Soler-Soler J. Clinical clues to the causes of large pericardial effusions. *Am J Med*. 2000;109:95-101.
6. Nagueh SF, Appleton CP, Gillebert TC, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography. *J Am Soc Echocardiogr*. 2009;22:107-133.

Global Public Health
Vol. 6, No. 6, September 2011, 669–677

Addressing the burden of post-conflict surgical disease – Strategies from the North Caucasus

Karsten Lunze[a]* and Fatima I. Lunze[b,c]

[a]*Preventive Medicine, Boston University, Boston, MA, USA;* [b]*Children's Hospital Boston, Boston, MA, USA;* [c]*North-Ossetian State Medical Academy, Vladikavkaz, Republic of North Ossetia-Alania, Russian Federation*

(*Received 24 June 2010; final version received 29 November 2010*)

The 2004 terror attack on a school in Beslan, North Caucasus, with more than 1300 children and their families taken hostage and 334 people killed, ended after extreme violence. Following the disaster, many survivors with blast ear injuries developed complications because no microsurgery services were available in the region. Here, we present our strategies in North Ossetia to strengthen subspecialty surgical care in a region of instable security conditions.

Disaster modifies disease burden in an environment of conflict-related health-care limitations. We built on available secondary care and partnered international with local stakeholders to reach and treat victims of a humanitarian disaster. A strategy of mutual commitment resulted in treatment of all consenting Beslan victims with blast trauma sequelae and of non disaster-related patients.

Credible, sustained partnerships and needs assessments beyond the immediate phases after a disaster are essential to facilitate a meaningful transition from humanitarian aid to capacity building exceeding existing insufficient standards. Psychosocial impacts of disaster might constitute a barrier to care and need to be assessed when responding to the burden of surgical disease in conflict or post-conflict settings. Involving local citizen groups in the planning process can be useful to identify and access vulnerable populations. Integration of our strategy into broader efforts might strengthen the local health system through management and leadership.

Keywords: burden of surgical disease; post-conflict; capacity building; North Caucasus; Beslan

Background

The North Caucasus remains a region of frequent human rights violations and resurging violence from armed opposition groups (Lunze 2009). The recent conflict over South Ossetia and ongoing attacks from militant groups in Chechnya, Ingushetia and Dagestan illustrate its political instability. Most non-governmental organisations (NGOs) have withdrawn due to security concerns. Health systems in the region, economically most disadvantaged within the Russian Federation, are weakened from past conflicts; while they provide primary care, they lack the capacity for specialised services.

On 1 September 2004, a group of terrorists attacked a school in Beslan, a small town with a population of 30,000, situated in the Republic of North Ossetia-Alania

*Corresponding author. Email: karsten.lunze@post.harvard.edu

ISSN 1744-1692 print/ISSN 1744-1706 online

DOI: 10.1080/17441692.2011.557667
http://www.informaworld.com

(an autonomous republic that is a part of the Russian Federation and neighbours Chechnya). In Russia, 1 September is traditionally the first day of school after summer vacation, when pupils together with their families gather in schools to celebrate. At the Beslan school, the terrorist group took more than 1300 children and their families hostage. After 3 days of what is now considered one of the worst massacres on a civilian population in Europe in recent history (Moscardino *et al*. 2010), Russian security forces stormed the building and ended the siege with the use of heavy artillery. The incident left 334 victims dead and many more injured. During the final storm, indoor bomb explosions in the gymnasium caused blast injuries in numerous survivors. Among survivors, traumatic ear defects are the most common blast injury (DePalma *et al*. 2005), which are amenable to relatively straightforward treatment but require subspecialty management (Wolf *et al*. 2009).

In peripheral regions of the Russian Federation, patients requiring services beyond local capacity are referred to central institutions following a quota system which provides a limited number of grants for specialized services (WHO 2005). While initially complex traumatic injuries resulting from the Beslan terror act had been addressed in North Ossetian hospitals or at major centres elsewhere within the Russian Federation (Schreeb *et al*. 2004), several years later many victims had still not received comprehensive medical care beyond the acute immediate response. Traumatic blast ear defects resulting from bomb explosions need to be assessed promptly after the incident. Delays in treatment are associated with further damages (Wolf *et al*. 2009), which may require extensive microsurgical reconstruction and expertise typically found only at major developed medical centres.

Two years after the disaster, clinicians at the Vladikavkaz Children's Hospital in the capital of North Ossetia-Alania observed an increased burden of post-traumatic middle ear injury. Tympanic lesions left untreated caused different pathological reactions in the mucosal and bony structures of the middle ear, requiring extensive microsurgical reconstruction procedures to limit destructive processes. In some cases middle ear prostheses were needed in order to guarantee an effective conduction of acoustic waves in the tympanon, to attempt an improvement in hearing or to prevent further hearing loss. No surgical capacity was available in or near North Ossetia to perform the necessary microsurgery for these patients.

Following an appeal from the Vladikavkaz Children's Hospital directed at the last author of this article, who trained there, she procured an ear-nose-throat (ENT) operation microscope and shipped it to Vladikavkaz Children's Hospital. Due to lack of local surgical subspecialty expertise, this was insufficient to meet the Beslan victims' needs. Health facilities in North Ossetia are outdated and poorly equipped. Personnel are skilled in general surgical services, but lack training and dedicated equipment for microsurgical procedures; and available services are poorly coordinated due to insufficiencies in management and leadership in the health sector.

Methods

The two authors initially arranged for consultation and treatment in Germany for patients from the Caucasus region with complex middle ear injuries. However, this approach soon proved unaffordable and unsustainable due to high transport and medical care costs. Most importantly, treatment abroad did not address the unavailability of microsurgical services in the North Caucasus. The two authors

therefore partnered with two ENT surgeons skilled in otologic operation techniques (Professor Thomas Eichhorn, Cottbus, and Dr Christian Offergeld, Freiburg; both in Germany), the Vladikavkaz Children's Hospital, the North Ossetian Ministry of Health, and local citizen groups in order to reassess the situation 2 years after the terror act and to formulate the following objectives, plans and strategies:

(1) Partner local government authorities, hospital faculty and staff and academic institutions as well as citizen groups, to identify and address the immediate need for surgical ENT treatment for Beslan victims and to offer them comprehensive treatment.
(2) Explore management and leadership challenges that led to the current gap.
(3) Create capacity by appropriately equipping and training surgeons from the region in microsurgical techniques with the long-term goal to establish comprehensive subspecialty services in the region.

Building on available structures, from the initial phase on, tasks such as strategic and administrative procedures, access to patients, patient care and follow-up activities were equally shared between international and local health professionals. This common approach allowed for clarifying goals and expectations, and identified opportunities for management and leadership improvement. It also helped ease procedural hurdles such as necessary formalities, accreditations and required permits, and allowed us to operate freely in a highly politicised environment where security concerns limit the operability of many organisations. In order to assess the local context from a supply and demand perspective, we conducted an assessment of local resources, infrastructure and surgical needs. Medical faculty and citizen groups in North Ossetia delivered the necessary data.

Results

While dedicated operation room capacity including anesthesia and basic surgical supplies existed and surgical care is established in North Ossetia, there was an almost complete lack of supplies and equipment for specialised surgery and microsurgery (see Table 1). Following the determination and coordination of available resources in

Table 1. Available and needed resources for specialised surgery at the Children's Hospital Vladikavkaz, North Ossetia-Alania.

Available resources	Local needs
Anesthesia machines and gas supplies	Operation microscope (with observer tube for teaching purposes), sterile covers and replacement lamps
Sterilising equipment	Electrocouter with ground plates and cables
Surgical gowns, caps, masks, gloves and drapes	Complete sets of dedicated instruments for ENT microsurgery
Elastic bandages, swabs and dressings	Microsurgical scalpel blades
Normal saline and Ringer's solutions	Absorbable haemostatic sponges
Needles, syringes	Dedicated suture material
Catecholamines	Dedicated drainage catheters
Antibiotics	
Disinfectants	

North Ossetia, we procured further specialised supplies and equipment for microsurgery to complement existing material. To incorporate best surgical practices into local care, one of the ENT specialists (Professor Thomas Eichhorn) at his institution in Germany trained a North Ossetian surgeon (Dr Zemfira Tsorieva) in microsurgical skills, who became competent to identify suitable patients, coordinated paediatric and adult surgical as well as anaesthesiologic services available in North Ossetia and ensured follow-up of patients in the post-operative phase.

Victims were identified and characterised using data from medical faculty and citizen groups in North Ossetia, as well as international academic and WHO sources (Schreeb *et al*. 2004). This assessment of specialised surgical needs yielded 19 patients with complex ear pathologies (see Table 2). Since many victims were mentally traumatised (Parfitt 2004), we consulted with the victims' representatives and human rights groups to assist medical staff at the Children's Hospital Vladikavkaz in accessing eligible patients. As a result, 10 identified victims were evaluated for surgical interventions by locally trained staff, who also obtained written informed consent from 14 eligible patients (six of whom were victims of the Beslan disaster), provided preoperative care and planned for operation room capacities.

We were confronted with four victims for whom surgical treatment was indicated but who refused treatments for psychosocial reasons, consistent with similar accounts from citizen groups. We were unable to further characterise the psychosocial burden and mental disease among the Beslan victims or to quantify the number of victims who declined treatment for those reasons.

The Children's Hospital Vladikavkaz provided operation room management, anesthesia staff and equipment as well as nursing staff for both adult and paediatric patients. Our team, including international volunteers and local surgeons, performed and documented, in total, 15 comprehensive microsurgical operations mainly for complex middle ear pathologies, including one additional non-elective emergency procedure, without intra- or post-operative complications (see Table 3). During the operations, local adult and paediatric surgeons from the area were instructed in microsurgery techniques.

All patients received care at no cost to them and without informal payments. During our activities in North Ossetia, we operated unhindered, with support from the North Ossetian health minister and assisted by one of his staff members. Two German journalists video-documented our activities and reported on the reactions of the local population without restrictions.

Table 2. Needs assessment of patient recruitment for specialised otological care after the Beslan disaster.

	Number
Hostages held at school in Beslan	1355
Hostages killed	334
Victims hospitalised	661
Victims requiring intensive care	110
Victims initially identified with post-traumatic ear disease	140
Victims identified with post-traumatic chronic middle ear otitis after two years	19
Victims identified with indication for specialised surgery	6

Table 3. Patient characteristics, diagnoses and interventions at the Children's Hospital Vladikavkaz, North Ossetia-Alania.

Patient characteristics	
Median age (range)	15 years (1.5 months–44 years)
Gender	5 females
	10 males
Diagnoses	10 cases of post-traumatic tympanic perforation
	Four cases of chronic otitis media
	One case of acute mastoiditis
Interventions	Thirteen tympanoplasties (including six with reconstruction of ossicular chain, two with adenotomy, one with ossicular prosthesis, and one with ossicular prosthesis and mastoidectomy)
	One tympanic tube insertion
	One emergency mastoidectomy

Management and leadership opportunities were identified as the need for improvement in coordination of health services and their availability to vulnerable populations; for mobilisation of current human resource potential by training existing faculty and junior health professionals; as well as for extension of microsurgery capacities to other surgical specialties (e.g., ophthalmology) and outreach to neighbouring post-conflict regions, such as South Ossetia, Chechnya and Ingushetia.

Discussion

This partnership to address the burden of surgical disease in the North Caucasus region resulted from an act of violence and a humanitarian disaster. In conflict and post-conflict situations, the most vulnerable populations are most difficult to reach. Partnering international volunteers with a variety of local stakeholders and involving citizen groups, such as victims' representatives and human rights organisations, lent credibility to reaching out to victims of the Beslan disaster and treating all eligible consenting individuals with sequelae of blast injuries resulting from insufficient subspecialty services. Key stakeholders were the health ministry, which oversees all health-care related activities, and the medical academy, which bundles all medical training and postgraduate medical education.

Local clinicians instructed during this collaboration continue to provide subspecialty patient care and train other providers in peripheral facilities. Most importantly – based on process evaluations and clinical outcomes – they will shape future training activities, as effective capacity building in the surgical specialties will require a strong commitment to education (Lancet Editorial 2010). Thus, the conjoint strategy started to address the local burden of surgical disease by strengthening subspecialty services for the region. Adequate, sustainable secondary level care, not only in acute emergency responses but also in longer-term post-conflict contexts and adapted to local needs, is fundamental for effective health systems, but often overlooked (Campbell and Doull 2010).

Our concerted approach, built on outdated but existing structures of secondary care, involved local resources from the beginning. Middle-income countries such

as those in the former Soviet Union or South America offer particular opportunities to address surgical burden beyond general surgery. Unlike in most low-income countries, where appropriate anesthesia services are severely limited (Hodges *et al*. 2007), we could rely on effective anesthesia capacity in North Ossetia.

Our needs assessment found poor infrastructure, inadequate equipment and supplies, and health professionals – albeit sufficient in number – who were inadequately trained. These factors represent typical barriers to appropriate and effective delivery of surgical services (Spiegel and Gosselin 2007). Our findings are consistent with systematic surveys suggesting that strengthening of infrastructure, supplies and procedures in low- and midle-income countries is urgently needed (Kushner *et al*. 2010). Substandard facilities threaten patients' outcomes (Lancet Editorial 2010), and effective, safe surgery is no luxury for midle-income countries: although this has not yet been studied for subspeciality services, there is increasing evidence that the cost-effectiveness ratio of surgical services might compare favourably with selected primary health interventions (Debas *et al*. 2006).

An evaluation conducted immediately after the Beslan disaster concluded that early post-trauma emergency care for victims was appropriately handled by local and national health resources, whereas international assistance – that unlike in other emergencies the authorities of the Russian Federation had requested – was deemed excessive, inappropriate and largely ignoring local needs (Schreeb *et al*. 2004). Our own assessment years after the disaster found a disease burden which was the result of insufficient subspecialty services.

We therefore advocate for periodical, reliable data collection beyond the short- and mid-term phases after a disaster, particularly once international attention and media coverage have faded, to reveal how both needs and available resources develop over time and in changing political environments, and to facilitate a meaningful transition from necessary humanitarian aid to appropriate partnerships for development.

Rather than reflecting the mere availability of services, meaningful needs assessments have to distinguish whether conflicts increase or modify disease burden, and whether they limit the availability of or access to health services (Lunze 2009, Kushner *et al*. 2010). We believe that the Beslan disaster led not only to an increase in the disease burden as we describe it, but also to impaired victims' care seeking for mental health reasons.

During the terror act in Beslan, victims had to endure extreme violence for several days under inhumane conditions. The resulting psychological trauma is considered a quaternary pattern of injury (Wolf *et al*. 2009). Although national and international organisations responded early to mental trauma with psychosocial counselling and rehabilitation (Parfitt 2004, UNICEF 2004), we suspect post-traumatic stress disorder to substantially impair victims access to treatment even years after the trauma. However, our planning focused on the delivery of surgical care rather than addressing potential barriers to accessing this care. Investigating and addressing mental health effects of terrorism and violence is immensely difficult in the complex and chaotic setting during and after disasters (North and Pfefferbaum 2002). Although it would have been relevant, we did not have the capacity nor did we attempt to measure to what extent psychosocial impacts and mental trauma affected patients' access to elective, subspecialised surgical services.

On the basis of our needs assessment, we had planned operations for 19 identified patients with ear complications. In fact, only a minority of six Beslan victims

consented to an operation. The majority of patients were operated on for advanced pathologies less commonly encountered in effective health systems, which were not conflict-related. Even assuming that some families raised sufficient funds to access care elsewhere, we believe that a number of disaster victims did not reach our services due to psychosocial barriers. We consider the negative impact of terror on survivors' mental health, mediated even after years by daily stressors of a post-conflict society with ongoing violence (Miller and Rasmussen 2010), to have impaired our recruitment of this vulnerable patient population for surgical treatment.

Several studies investigating mental health in Beslan victims confirmed our anecdotal observations that the terror attack persistently impaired the psychological well-being of victims as well as of their families and caregivers (Scrimin *et al*. 2006, Moscardino *et al*. 2008). Regardless of being directly or indirectly exposed, the disaster influenced the reorganisation of family life and the disruption of community ties (Moscardino *et al*. 2010). Cultural values and gender differences factor into victims' coping strategies and are inherently complex in this society, where deeply rooted traditions shape everyday life (Moscardino *et al*. 2007).

Therefore, assessment of mental health effects, neglected during our own planning, should be part of programming efforts when responding to the burden of surgical disease, particularly in conflict or post-conflict settings. This could be done in collaboration with groups or organisations with expertise in post-conflict psychosocial health who have an established relation with the population. Involving human rights groups in the planning process, albeit a delicate step, can assist in identifying and accessing these populations, to which international organisations have less access, for security and various other reasons.

Limitations

Adequate funding is a crucial requirement to transition the response to the global burden of surgical disease from – in many cases – helpful short-term volunteer surgical missions to sustainable and more meaningful efforts (Farmer and Kim 2008, Farmer 2010). Given the current working conditions for NGOs in the Russian Federation, in spite of uniting a whole variety of stakeholders, we deliberately chose to act as members of civil society and not as an organisation, in order to safeguard our own security and minimise risks for the organisations we worked with. This has severely limited our ability to seek funding, which is difficult to obtain for a region with a volatile security situation and travel restrictions for foreign personnel.

More long-term efforts than our interventions are required to ensure ongoing appropriate surgical care delivery at international standards. To achieve proficiency in specialised techniques such as otologic operations usually takes several years of postgraduate training at a dedicated institution. While this is hardly feasible in a middle-income country, training health personnel abroad carries the risk of brain drain through those who are not willing or able to return to their home country.

Conclusions

We identified a number of strategies that we believe might be helpful when planning capacity building for surgical care in post-conflict settings:

(1) Building on available secondary care in midle-income countries can address the post-conflict burden of surgical disease beyond general surgery.
(2) Partnering international with local stakeholders, including citizen groups, can create credible partnerships to access vulnerable populations in politicised environments.
(3) Needs assessments beyond the short- and mid-term phases after a disaster are essential to facilitate a meaningful transition from humanitarian aid to partnerships for development.
(4) Psychosocial impacts might affect both needs and care seeking and should be assessed as part of a comprehensive approach when responding to the burden of surgical disease in conflict or post-conflict settings.
(5) Health-care delivery planning will have to accommodate a case mix that will not only include conflict-related burdens, but also advanced and natural course pathologies resulting from health systems insufficiencies.

Given how rapidly post-conflict situations change, these strategies will have to be adapted over time and place to given – and changing – needs, political and security circumstances. Short-term surgical missions focusing on a limited range of pathologies have a recognised value and have made substantial contributions to many of the disadvantaged in this world (Farmer and Kim 2008). In a public health framework, humanitarian operations and skills training have been considered selective preventive interventions of political violence at the level of society at large (De Jong 2010). However, sustainable change requires sustained investments of time and resources beyond an initially vertical mission and critical analysis, in order to create the conditions that incentivise current health professionals to continue working in and developing their professional environment. We see a true value of our strategies in the ongoing commitment to our partnership, which now aims at integrating the important pillar of high-quality surgical care delivery into broader efforts of strengthening the local health system through management and leadership.

References

Campbell, F. and Doull, L., 2010. Provision of secondary care in fragile state contexts. *Lancet*, 375 (9711), 260–261.

Debas, H.T., Gosselin, R., McCord, C., and Thind, A., 2006. Surgery. *Disease control priorities in developing countries*. 2nd ed. New York: Oxford University Press, 1245–1260.

De Jong, J.T., 2010. A public health framework to translate risk factors related to political violence and war into multi-level preventive interventions. *Social Science and Medicine*, 70 (1), 71–79.

DePalma, R.G., Burris, D.G., Champion, H.R., and Hodgson, M.J., 2005. Blast injuries. *New England Journal of Medicine*, 352 (13), 1335–1342.

Farmer, D.L., 2010. The need for sustainability in contemporary global health efforts: missions vs. mission. *Archives of Surgery*, 145 (8), 752–753.

Farmer, P.E. and Kim, J.Y., 2008. Surgery and global health: a view from beyond the OR. *World Journal of Surgery*, 32 (4), 533–536.

Hodges, S.C., Mijumbi, C., Okello, M., McCormick, B.A., Walker, I.A., and Wilson, I.H., 2007. Anaesthesia services in developing countries: defining the problems. *Anaesthesia*, 62 (1), 4–11.

Kushner, A.L., Cherian, M.N., Noel, L., Spiegel, D.A., Groth, S., and Etienne, C., 2010. Addressing the Millennium Development Goals from a surgical perspective: essential

surgery and anesthesia in 8 low- and middle-income countries. *Archives of Surgery*, 145 (2), 154–159.
Lancet Editorial, 2010. What is the point of surgery? *Lancet*, 376 (9746), 1025.
Lunze, K., 2009. Health and human rights: no miracle in postconflict Chechnya. *Lancet*, 374 (9704), 1809–1810.
Miller, K.E. and Rasmussen, A., 2010. War exposure, daily stressors, and mental health in conflict and post-conflict settings: bridging the divide between trauma-focused and psychosocial frameworks. *Social Science and Medicine*, 70 (1), 7–16.
Moscardino, U., Axia, G., Scrimin, S., and Capello, F., 2007. Narratives from caregivers of children surviving the terrorist attack in Beslan: issues of health, culture, and resilience. *Social Science and Medicine*, 64 (8), 1776–1787.
Moscardino, U., Scrimin, S., Capello, F., Altoe, G., and Axia, G., 2008. Psychological adjustment of adolescents 18 months after the terrorist attack in Beslan, Russia: a cross-sectional study. *Journal of Clinical Psychiatry*, 69 (5), 854–859.
Moscardino, U., Scrimin, S., Capello, F., and Altoe, G., 2010. Social support, sense of community, collectivistic values, and depressive symptoms in adolescent survivors of the 2004 Beslan terrorist attack. *Social Science and Medicine*, 70 (1), 27–34.
North, C.S. and Pfefferbaum, B., 2002. Research on the mental health effects of terrorism. *Journal of the American Medical Association*, 288 (5), 633–636.
Parfitt, T., 2004. How Beslan's children are learning to cope. *Lancet*, 364 (9450), 2009–2010.
Schreeb, J.V., Kalmykov, A., Riddez, L., and Rosling, H., 2004. Emergency care following the terrorist attack in Beslan, North Ossetia, Russian Federation, 2004. *International Journal of Disaster Medicine*, 2 (1&2), 41–47
Scrimin, S., Axia, G., Capello, F., Moscardino, U., Steinberg, A.M., and Pynoos, R.S., 2006. Posttraumatic reactions among injured children and their caregivers 3 months after the terrorist attack in Beslan. *Psychiatry Research*, 141 (3), 333–336.
Spiegel, D.A. and Gosselin, R.A., 2007. Surgical services in low-income and middle-income countries. *Lancet*, 370 (9592), 1013–1015.
UNICEF, 2004. *Children in the Russian Federation*. Available from: http://www.unicef.org/ceecis/BackgroundRussFed2004.doc [Accessed 12 October 2010].
WHO, 2005. *Health Action in the North Caucasus*. Available from: http://www.who.int/hac/crises/rus/sitreps/Health%20Action%20Feb_Mar_2005%20%20.pdf [Accessed 22 February 2011].
Wolf, S.J., Bebarta, V.S., Bonnett, C.J., Pons, P.T., and Cantrill, S.V., 2009. Blast injuries. *Lancet*, 374 (9687), 405–415.

General Internal Medicine
Boston University School of Medicine
2011 Publications – M-Z

1. Marrast L, Evans M, Ozonoff A, **Henault L**, **Rose AJ**. Using highly variable warfarin dosing to identify patients at risk for adverse events. Thrombosis Journal. 2011 Oct;9(1):14. PMID: 21985504.

2. Matsuyama RK, Wilson-Genderson M, Kuhn L, Moghanaki D, Vachhani H, **Paasche-Orlow M**. Education level, not health literacy, associated with information needs for patients with cancer. Patient Education and Counseling. 2011 Dec;85(3):e229-36. PMID: 21543182.

3. McCloskey L, Condon R, **Shanahan CW**, Wolff J, Culler C, **Kalish R**. Public health, medicine, and dentistry as partners in community health: A pioneering initiative in interprofessional, practice-based education. J Public Health Manag Pract. 2011 Jul-Aug;17(4):298-307. PMID: 21617403.

4. McCormack LA, Rush SR, Kandula NR, **Paasche-Orlow MK**. Health literacy research: Looking forward. J Health Commun. 2011 Sep;16 Suppl 3:5-8. PMID: 21951238.

5. McInnes DK, Solomon JL, Bokhour BG, Asch SM, Ross D, Nazi KM, **Gifford AL**. Use of electronic personal health record systems to encourage HIV screening: An exploratory study of patient and provider perspectives. BMC Res Notes. 2011 Aug;4(1):295. PMID: 21843313.

6. **Medow MA**, Lucey CR. A qualitative approach to Bayes' theorem. Evid Based Med. 2011 Dec;16(6):163-7. PMID: 21862499.

7. Meltzer EC, Rybin D, **Saitz R**, **Samet JH**, Schwartz SL, Butler SF, **Liebschutz JM**. Identifying prescription opioid use disorder in primary care: Diagnostic characteristics of the Current Opioid Misuse Measure (COMM). Pain. 2011 Feb;152(2):397-402. PMID: 21177035. PMCID: PMC3027065.

8. Merrick ES, Hodgkin D, Garnick DW, Horgan CM, Panas L, Ryan M, Blow FC, **Saitz R**. Older adults' inpatient and emergency department utilization for ambulatory-care-sensitive conditions: Relationship with alcohol consumption. J Aging Health. 2011 Feb;23(1):86-111. PMID: 20935248.

9. Myers JS, Gojraty S, Yang W, **Linsky A**, Airan-Javia S, Polmano RC. A randomized-controlled trial of computerized alerts to reduce unapproved medication abbreviation use. J Am Med Inform Assoc. 2011 Jan-Feb;18(1):17-23. PMID: 21131606. PMCID: PMC3005872.

10. **Naimi TS**. "Gray area" alcohol consumption and the U.S. Dietary Guidelines: A comment on Dawson and Grant (2011). J Stud Alcohol Drugs. 2011 Jul;72(4):687. PMID: 21683052.

11. **Naimi TS**. The cost of alcohol and its corresponding taxes in the U.S.: A massive public subsidy of excessive drinking and alcohol industries. Am J Pre Med. 2011;41(5):546-7. PMID: 22011428.

12. Ohlsson C, Wallaschofski H, Lunetta KL, Stolk L, Perry JR, Koster A, Petersen AK, Eriksson J, Lehtimäki T, Huhtaniemi IT, Hammond GL, Maggio M, Coviello AD; EMAS Study Group, Ferrucci L, Heier M, Hofman A, Holliday KL, Jansson JO, Kähönen M, Karasik D, Karlsson MK, Kiel DP, Liu Y, Ljunggren O, Lorentzon M, Lyytikäinen LP, Meitinger T, Mellström D, Melzer D, Miljkovic I, Nauck M, Nilsson M, Penninx B, Pye SR, Vasan RS, Reincke M, Rivadeneira F, Tajar A, Teumer A, Uitterlinden AG, Ulloor J, Viikari J, Völker U, Völzke H, Wichmann HE, Wu TS, Zhuang WV, Ziv E, Wu FC, Raitakari O, Eriksson A, Bidlingmaier M, Harris TB, Murray A, de Jong FH, **Murabito JM**, Bhasin S, Vandenput L, Haring R. Genetic determinants of serum testosterone concentrations in men. PLoS Genet. 2011 Oct;7(10):e1002313. Epub 2011 Oct 6. PMID: 21998597.

13. Orwat J, **Saitz R**, Tompkins CP, **Cheng DM**, Dentato MP, **Samet JH**. Substance abuse treatment utilization among adults living with HIV/AIDS and alcohol or drug problems. J Subst Abuse Treat. 2011 Jan;113(2-3):165-71. PMID: 21700412.

14. Orwat J, **Samet JH**, Tompkins CP, **Cheng DM**, Dentato MP, **Saitz R**. Factors associated with attendance in 12-step groups (Alcoholics Anonymous/Narcotics Anonymous) among adults with alcohol problems living with HIV/AIDS. Drug Alcohol Depend. 2011 Jan;113(2-3):165-71. PMID: 20832197. doi:10.1016/j.drugalcdep.2010.07.021.

15. Osborn CY, **Paasche-Orlow MK**, Bailey SC, Wolf MS. The mechanisms linking health literacy to behavior and health status. Am J Health Behav. 2011 Jan-Feb;35(1):118-28. PMID: 20950164. PMCID: PMC3085858.

16. **Paasche-Orlow MK**. Caring for patients with limited health literacy: A 76-year-old with multiple medical problems. JAMA. 2011 Sept;306(10):1122-29.

17. **Paasche-Orlow MK**, Hung L. Crossword puzzle: Curl up with a good one. J Health Commun. 2011 Sept;16 Suppl 3:328-9. PMID: 21951262.

18. Palfai TP, Zisserson R, **Saitz R**. Using personalized feedback to reduce alcohol use among hazardous drinking college students: The moderating effect of alcohol-related negative consequences. Addict Behav. 2011 May;36(5):539-42. PMID: 21295919.

19. **Phillips CE**, Rothstein JD, Beaver K, **Sherman BJ**, **Freund KM**, **Battaglia TA**. Patient Navigation to increase mammography screening among inner city women. JGIM. 2011 Feb;26(2):123-9. PMID: 20931294. doi: 10.1007/s11606-010-1527-2.

20. Piccini JP, Fraulo ES, Ansell JE, Fonarow GC, Gersh BJ, Go AS, **Hylek EM**, Kowey PR, Mahaffey KW, Thomas LE, Kong MH, Lopes RD, Mills RM, Peterson ED. Outcomes registry for better informed treatment of atrial fibrillation: Rationale and design of ORBIT-AF. Am Heart J. 2011 Oct;162(4):606-612.e1. PMID: 21982650.

21. Pugh MJ, **Berlowitz DR**, Rao JK, Shapiro G, Avetisyan R, **Hanchate A**, Jarrett K, Tabares J, Kazis LE. The quality of care for adults with epilepsy: An initial glimpse using the QUIET measure. BMC Health Serv Res. 2011 Jan;11:1-9. PMID: 21199575.

22. Pugh MJ, Hanlon JT, Wang CP, Semla T, Burk M, Amuan ME, Lowery A, Good CB, **Berlowitz DR**. Trends in use of high-risk medications for older veterans: 2004 to 2006. J Am Geriatr Soc. 2011 Aug 30. doi: 10.1111/j.1532-5415.2011.03559.x. PMID: 21883108.

23. Pugh MJ, Starner CI, Amuan ME, **Berlowitz DR**, Horton M, Marcum ZA, Hanlon JT. Exposure to potentially harmful drug-disease interactions in older community-dwelling veterans based on the healthcare effectiveness data and information set quality measure: Who is at risk? J Am Geriatr Soc. 2011;59(9):1673-78. doi: 10.1111/j.1532-5415.2011.03524.x. PMID: 21831166.

24. Pyne JM, Fortney JC, Curran GM, Tripathi S, Atkinson JH, Kilbourne AM, Hagedorn HJ, Rimland D, Rodriguez-Barradas MC, Monson T, Bottonari KA, Asch SM, **Gifford AL**. Effectiveness of collaborative care for depression in HIV clinics. Arch Intern Med. 2011 Jan;171(1):23-31.PMID: 2122065.

25. Raj A, Saggurti N, **Cheng DM**, Dasgupta A, **Bridden C**, Pradeshi M, **Samet JH**. Transactional sex risk and STI among HIV-infected female sex workers and HIV-infected male clients of FSWs in India. AIDS Care. 2011 Nov;23(11):1374-81. PMID: 21711167.

26. Rammohan V, Hahn RA, Elder R, Brewer R, Fielding J, **Naimi TS**, Toomey TL, Chattopadhyay SK, Zometa C; Task Force on Community Preventive Services. Effects of dram shop liability and enhanced overservice law enforcement initiatives on excessive alcohol consumption and related harms two community guide systematic reviews. Am J Prev Med. 2011 Sep;41(3):334-43. PMID: 21855749.

27. Reif S, Larson MJ, **Cheng DM**, **Allensworth-Davies D**, **Samet J**, **Saitz R**. Chronic disease and recent addiction treatment utilization among alcohol and drug dependent adults. Subst Abuse Treat Prev Policy. 2011 Oct;6:28. PMID: 22008255. PMCID: PMC3220629.

28. **Rose AJ**, **Berlowitz DR**, Ash AS, Ozonoff A, **Hylek EM**, Goldhaber-Fiebert JD. The business case for quality improvement: Oral anticoagulation for atrial fibrillation. Circ Cardiovasc Qual Outcomes. 2011 Jul;4(4):416-24. PMID: 21712521.

29. **Rose AJ**, Glickman ME, **D'Amore MM**, Orner MB, **Berlowitz D**, **Kressin NR**. Effects of daily adherence to antihypertensive medication on blood pressure control. J Clin Hypertens (Greenwich). 2011 Jun;13(6):416-21. doi: 10.1111/j.1751-7176.2011.00427. PMID: 21649841. PMCID: PMC3111950.

30. **Rose AJ**, **Hylek EM**, **Berlowitz DR**, Ash AS, Reisman JI, Ozonoff A. Prompt repeat testing after out-of-range INR values: a quality indicator for anticoagulation care. Circ Cardiovasc Qual Outcomes. 2011 May;4(3):276-82. PMID: 21505156.

31. **Rose AJ**, **Hylek EM**, Ozonoff A, Ash AS, Reisman JI, **Berlowitz DR**. Risk-adjusted percent time in therapeutic range as a quality indicator for outpatient oral anticoagulation: Results of the Veterans AffaiRs study to Improve Anticoagulation (VARIA). Circ Cardiovasc Qual Outcomes. 2011 Jan;4(1):22-29. PMID: 21098779.

32. **Rose AJ**, **Hylek EM**, Ozonoff A, Ash AS, Reisman JI, Callahan PP, Gordon MM, **Berlowitz DR**. Relevance of current guidelines for organizing an anticoagulation clinic. Am J Manag Care. 2011 Apr;17(4):284-9. PMID: 21615198.

33. **Rose AJ**, Ozonoff A, **Berlowitz DR**, Ash AS, Reisman JI, **Hylek EM**. Reexamining the recommended follow-up interval after an in-range INR value: Results from the Veterans AffaiRs study to Improve Anticoagulation (VARIA). Chest. 2011 Aug;140(2):359-65. PMID: 21310837.

34. Rush SR, **Paasche-Orlow MK**. Incorporating health literacy into larger operational environments. J Health Commun. 2011 Sept;16 Suppl 3:9-10. PMID: 21951239.

35. Russell SL, Katz RV, Wang MQ, Lee R, Green BL, **Kressin NR**, Claudio C. Belief in AIDS origin conspiracy theory and willingness to participate in biomedical research studies: Findings in whites, blacks, and hispanics in seven cities across two surveys. HIV Clin Trials. 2011 Jan-Feb;12(1):37-47. PMID: 21388939.

36. **Saitz R**. Worldwide evidence-based medicine activities. Evid Based Med. 2011 Feb;16(1):5. doi:10.1136/ebm1172. PMID: 21228051.

37. **Samet JH**. Russia and human immunodeficiency virus-beyond crime and punishment. Addiction. 2011 Nov;106(11):1883-5. doi: 10.1111/j.1360-0443.2011.03523.x. PMID: 21851440.

38. Schroy PC 3rd, Emmons K, Peters E, Glick JT, Robinson PA, Lydotes MA, Mylvanaman S, Evans S, Chaisson C, Pignone M, Prout M, **Davidson P**, Heeren TC. The impact of a novel computer-based decision aid on shared decision making for the colorectal cancer screening: a randomized trial. Med Decis Making. 2011 Jan-Feb;31(1):93-107. PMID: 20484090.

39. Siegel M, DeJong W, **Naimi TS**, Heeren T, Rosenbloom DL, Ross C, Ostroff J, Jernigan DH. Alcohol brand preferences of underage youth: Results from a pilot study among a national sample. Subst Abus. 2011 Oct;32(4):191-201.

40. Siegel MB, **Naimi TS**, Cremeens JL, Nelson DE. Alcoholic beverage preferences and associated drinking patterns and risk behaviors among high school youth. Am J Prev Med. 2011 Apr;40(4):419-26. PMID: 21406275.

41. Silverman JG, Decker MR, **Cheng DM**, Wirth K, Saggurti N, McCauley HL, Falb KL, Donta B, Raj A. Gender-based disparities in infant and child mortality based on maternal exposure to spousal violence: The heavy burden borne by Indian girls. Arch Pediatr Adolesc Med. 2011 Jan;165(1):22-7. PMID: 21199976.

42. Silverman JG, Raj A, **Cheng DM**, Decker MR, Coleman S, **Bridden C**, Pardeshi M, Saggurti N, **Samet JH**. Sex Trafficking and Initiation-Related Violence, Alcohol Use, and HIV Risk Among HIV-Infected Female Sex Workers in Mumbai, India. J Infect Dis. 2011 Dec;204 Suppl 5:S1229-34. PMID: 22043037.

43. **Somwaru LL**, Arnold AM, Cappola AR. Predictors of thyroid hormone initiation in older adults: Results from the Cardiovasular Health Study. J Gerontol A Biol Sci Med Sci. 2011 July;66A(7):809-14. doi: 10.1093/gerona/glr063. PMID: 21628677.

44. Southern WN, Drainoni ML, Smith BD, Christiansen CL, McKee D, **Gifford AL**, Weinbaum CM, Thompson D, Koppelman E, Maher S, Litwin AH. Hepatitis C testing practices and prevalence in a high-risk urban ambulatory care setting. J Viral Hepat. 2011 Jul;18(7):474-81. PMID: 20497311. PMCID: PMC2990787. doi: 10.1111/j.1365-2893.2010.01327.x.

45. Tsao CW, Gona P, Salton C, **Murabito JM**, Oyama N, Danias PG, O'Donnell CJ, Manning WJ, Yeon SB. Relationship between central and peripheral atherosclerosis and left ventricular dysfunction in a community population. Vasc Med. 2011 Aug;16(4):253-9. PMID: 21708875.

46. **Tsui JI**, Herman DS, Kettavong M, Anderson BJ, Stein MD. Chronic pain and hepatitis C virus infection in opioid dependent injection drug users. J Addict Dis. 2011 Apr;30(2):91-7. PMID: 21491290. PMCID: PMC3089758.

47. **Tsui JI**, Herman DS, Kettavong M, Anderson BJ, Stein MD. Escitalopram is associated with reductions in pain severity and pain interference in opioid dependent patients with depressive symptoms. Pain. 2011 Nov;152(11):2640-4. PMID: 21924552.

48. Vimalananda VG, Rosenzweig JL, Cabral HJ, **David MM**, **Lasser KE**. Comparison of diabetes control among Haitians, African Americans and non-Hispanic white in an urban safety-net hospital. Diabetes Care. 2011 Jan;34(1):58-60.

49. Volandes AE, Ferguson LA, Davis AD, Hull NC, Green MJ, Chang Y, Deep K, **Paasche-Orlow MK**. Assessing end-of-life preferences for advanced dementia in rural patients using an educational video: A randomized controlled trial. J Palliat Med. 2011 Feb;14(2):169-77. PMID: 21254815.

50. Walter S, Atzmon G, Demerath EW, Garcia ME, Kaplan RC, Kumari M, Lunetta KL, Milaneschi Y, Tanaka T, Tranah GJ, Völker U, Yu L, Arnold A, Benjamin EJ, Biffar R, Buchman AS, Boerwinkle E, Couper D, De Jager PL, Evans DA, Harris TB, Hoffmann W, Hofman A, Karasik D, Kiel DP, Kocher T, Kuningas M, Launer LJ, Lohman KK, Lutsey PL, Mackenbach J, Marciante K, Psaty BM, Reiman EM, Rotter JI, Seshadri S, Shardell MD, Smith AV, van Duijn C, Walston J, Zillikens MC, Bandinelli S, Baumeister SE, Bennett DA, Ferrucci L, Gudnason V, Kivimaki M, Liu Y, **Murabito JM**, Newman AB, Tiemeier H, Franceschini N. A genome-wide association study of aging. Neurobiol Aging. 2011 Nov;32(11):2109.e15-28. PMID: 21782286.

51. Wharam JF, Frank MB, Rosland A-M, **Paasche-Orlow MK**, Farber NJ, Sinsky C, Rucker L, Rask KJ, Barry MJ, Figaro K. "Pay-for-performance" as a quality improvement tool: Perceptions and policy recommendations of physicians and program leaders. Q Manage Health Care. 2011 July/Sept;20(3):234-45. PMID: 21725221.

52. Wilkinson JE, Lauer E, **Freund KM**, Rosen AK. Determinants of mammography in women with intellectual disabilities. J Am Board Fam Med. 2011 Nov;24(6):693-703. PMID: 22086812.

53. Xu W, Perera S, Medich D, Fiorito G, Wagner J, **Berger LK**, Greenspan SL. Height loss, vertebral fractures, and the misclassification of osteoporosis. Bone. 2011 Feb;48(2):307-11. PMID: 20870048.

54. Yu, T, Seger DL, **Lasser KE**, Karson AS, Fiskio JM, Seger AC, Bates DW. Impact of implementing alerts about medication black-box warnings in electronic health records. Pharmacoepidemiology and Drug Safety. 2011 Feb;20(2):192-202. PMID: 21254291. doi: 10.1002/pds.2088.

55. Zhai G, Teumer A, Stolk L, Perry JR, Vandenput L, Coviello AD, Koster A, Bell JT, Bhasin S, Eriksson J, Eriksson A, Ernst F, Ferrucci L, Frayling TM, Glass D, Grundberg E, Haring R, Hedman AK, Hofman A, Kiel DP, Kroemer HK, Liu Y, Lunetta KL, Maggio M, Lorentzon M, Mangino M, Melzer D, Miljkovic I; MuTHER Consortium, Nica A, Penninx BW, Vasan RS, Rivadeneira F, Small KS, Soranzo N, Uitterlinden AG, Völzke H, Wilson SG, Xi L, Zhuang WV, Harris TB, **Murabito JM**, Ohlsson C, Murray A, de Jong FH, Spector TD, Wallaschofski H. Eight common genetic variants associated with serum DHEAS levels suggest a key role in ageing mechanisms. PLoS Genet. 2011 Apr;7(4):e1002025. PMID: 21533175. PMCID: PMC3077384.

THROMBOSIS JOURNAL

ORIGINAL CLINICAL INVESTIGATION Open Access

Using highly variable warfarin dosing to identify patients at risk for adverse events

Lyndonna Marrast[1], Mary Evans[1], Al Ozonoff[2], Lori E Henault[1] and Adam J Rose[3,1*]

Abstract

Background: Patients who receive highly variable doses of warfarin may be at risk for poor anticoagulation control and adverse events. However, we lack a system to identify patients with the highest dose variability. Our objectives were to develop a scoring system to identify patients with high dose variability, and to validate this new measure by demonstrating that patients so identified have poor anticoagulation control and higher rates of adverse events (criterion validity).

Methods: We used a database of over 4, 000 patients who received oral anticoagulation in community practice between 2000-2002. We reviewed the charts of 168 patients with large warfarin dose variation and agreed on 18 risk factor definitions for high dose variability. We identified 109 patients with the highest dose variability (cases), as measured by coefficient of variation (CoV, SD/mean). We matched each case to two controls with low dose variability. Then, we examined all 327 charts, blinded to case/control status, to identify the presence or absence of the 18 risk factors for dose variability. We performed a multivariable analysis to identify independent predictors of high CoV. We also compared anticoagulation control, as measured by percent time in therapeutic range (TTR), and rates of adverse events between groups.

Results: CoV corresponded with other measures of anticoagulation control. TTR was 53% among cases and 79% among controls ($p < 0.001$). CoV also predicted adverse events. Six cases experienced a major hemorrhage versus 1 control ($p < 0.001$) and 3 cases had a thromboembolic event versus 0 control patients ($p = 0.04$). Independent predictors of high dose variability included hospitalization (OR = 21.3), decreased oral intake (OR = 12.2), use of systemic steroids (OR = 6.1), acetaminophen (OR = 4.0) and antibiotics (OR = 2.7; $p < 0.05$ for all).

Conclusion: CoV can be used to identify patients at risk for poor anticoagulation control and adverse events. This new measure has the potential to identify patients at high risk before they suffer adverse events.

Keywords: anticoagulants, dose variability, medication therapy management, risk factors, warfarin.

Background

Warfarin is the standard anticoagulation treatment for atrial fibrillation, venous thromboembolism (VTE), and mechanical heart valves [1-4]. Close monitoring of the International Normalized Ratio (INR) is required due to the drug's very narrow therapeutic window. Many factors can affect INR levels [1,5,6]. Values must be kept within range to reduce the risk of hemorrhage [7,8] and the risk of developing thromboembolism [9]. Previous studies have shown that patients experiencing better anticoagulation control have fewer such adverse events [10-14].

Assessment of adequate anticoagulation control has traditionally been determined by examining INR values themselves, through summary statistics such as percent time in therapeutic range (TTR) [15] or INR variability [16,17]. Several studies have explored the patient-level predictors of control as measured by TTR [10,18,19]. However, there is reason to believe that variability in warfarin doses could also serve to identify patients who are experiencing poorly controlled anticoagulation, thus placing them at risk for adverse events.

We therefore used a large, nationally representative database of community-based oral anticoagulation care

* Correspondence: adamrose@bu.edu
[3]Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, Bedford, MA, USA
Full list of author information is available at the end of the article

to address three related questions. First, we sought to develop a measure of warfarin dose variability that could be used to describe a population and identify patients with highly variable doses over time. Second, we sought to internally validate this new dose variability score as a measure of anticoagulation control using criterion validity. That is, we sought to demonstrate that patients identified as having high dose variability have worse anticoagulation control as measured by TTR and are at higher risk for adverse events than patients with less variability. Finally, through chart review, we sought to identify patient-level predictors of high dose variability. Our overarching goal was to develop a score that could be used to identify patients at high risk for complications.

Methods

Database

The Anticoagulation Consortium to Improve Outcomes Nationally (ACTION) study was a large prospective cohort study designed to assess the management of warfarin in community practice within the US [19-21]. A total of 101 participating sites in 31 states recruited 6761 patients receiving long-term oral anticoagulation. All sites used a freely-available software package called CoumaCare for tasks such as patient tracking and recording clinical data. In the database, clinicians updated patient's weekly warfarin dose at each visit. Because the present study relied upon chart reviews, we limited this study to the 47 sites of care that recorded complete notes for at least 90% of INR values. Excluded sites recorded notes only when the INR was not therapeutic. Therefore, this study was limited to 4489 patients.

Enrollment in ACTION occurred between April 2000 and February 2002. Patients were eligible to participate if they were 18 years or older and able to provide informed consent. All data were collected and their completeness rigorously ensured by McKesson HBOC, an independent data management organization. Missing data fields and data entry errors were resolved directly with the sites by the data coordinating center on a weekly basis before the data were transmitted to study investigators. The study protocol was approved by the Western Institutional Review Board of Olympia, WA, and by local review boards where they existed.

Patients were eligible for inclusion in the present study if they had an INR target range of 2-3. Indications for anticoagulation were grouped as follows: atrial fibrillation, venous thromboembolism, valvular heart disease/prosthetic heart valve, and all others. The database included demographics (age, gender, and race) and several comorbid conditions (coronary artery disease, congestive heart failure, hypertension, diabetes) as recorded by the patients' clinicians. Weekly dose of warfarin was recorded for all patients in the database, and was updated by clinicians at each visit. We used these weekly doses to assess the stability of warfarin dose over time for each patient, as will be explained below.

Chart Reviews

We performed two separate chart reviews, in our efforts to create a score that describes patients with high warfarin dose variability. The first review was implicit; it was performed by chart reviewers without relying upon pre-established definitions. Three physician examiners (LM, ME, and AJR) independently reviewed the charts of 168 patients who had a 2-fold or greater difference between the lowest and highest weekly warfarin dose (e.g. 14 mg/week versus 28 mg/week). The concept behind the review was to remain open to the possibilities of factors that may be present in the database rather than rely solely upon preconceived ideas. Next, the reviewers met and compiled a list of 18 variables believed to have played the greatest role in the dose variability. They reached a consensus regarding a standard definition for each variable in the chart review instrument (Table 1).

We found that the criterion used to identify patients with high dose variability (i.e. twofold or greater dose range) did not capture the dose variability we had in mind. Specifically, the method identified a relatively large proportion of patients with one or two outlier doses but otherwise stable dosing. Not all of the patients identified by this score seemed to be experiencing the highly variable anticoagulation control that we were trying to capture. We therefore decided to use the coefficient of variation (CoV) to characterize warfarin dose variability. CoV is defined as the standard deviation of the weekly warfarin dose divided by the mean weekly warfarin dose.

We labeled all patients with CoV greater than 0.2 as patients with high dose variation ("cases"). There were 123 such patients, representing 2.7% of the dataset. Patients with CoV below 0.05 (1019 patients, representing 23% of the dataset) were eligible to be controls. Each case was matched to 2 controls within the same site of care. Charts were excluded if: 1) there were no controls available to match the case patients or 2) the patient was new to warfarin (less than 1 month experience as of study entry). A total of 14 cases and 12 potential controls were removed for these reasons, leaving 109 cases and 218 controls.

The reviewers then independently reviewed charts to identify the 18 variables defined in the chart review instrument. This second review was explicit in that it relied upon the variable definitions described in the instrument. During this second chart review, reviewers were blinded to whether the patient was a case or a control patient. If a factor was present at any time, we recorded this indicator as "1" (present) versus "0" (not present). Each reviewer abstracted one-third of the charts. Fifty of the charts were

Table 1 Chart Review Instrument

	Variable	Definition
1	Diet	Any mention of "greens", specific foods high in vitamin K, and dietary content of vitamin K. DOES NOT INCLUDE statements that the vitamin K content of the diet is unchanged.
2	Dietary Supplements	Any mention of multivitamins, Ensure, Boost, Slimfast, etc. as they relate to vitamin K intake. DOES NOT INCLUDE simply listing a multivitamin in the medication list.
3	Adherence	Any mention of problems with adherence to pill-taking, including unauthorized self-adjustment of doses and memory issues. DOES NOT INCLUDE dose confusion after a hospital stay and DOES NOT INCLUDE aspects of adherence (diet, lab follow up, etc.) beyond pill-taking.
4	Hospital or Nursing Home Stay	Any mention of a hospital or nursing home stay EXCEPT for CHF (because that has its own variable - see below)
5	Nausea and Vomiting	Any mention
6	Decreased PO Intake or Decreased Appetite	Any mention
7	Diarrhea	Any mention
8	Decompensated CHF	Any mention of fluid overload, fluid retention, edema, pulmonary edema. Any titration of lasix doses, trending of weight regarding fluid status, use of metolazone (i.e. zaroxolyn), or any obvious CHF regimen. Any hospital admissions for fluid overload.
9	Alcohol	Any mention of alcohol except "denies." Exception - one serving per day or less does not count
10	Amiodarone	Any mention of amiodarone or its brand names "pacerone" or "cordarone."
11	Acetaminophen	Any mention of acetaminophen, products containing acetaminophen. Includes the abbreviation "APAP."
12	NSAIDS/COX-2 Inhibitors	Any mention at all, including mention in the medication list.
13	Procedures	Any mention of a procedure in conjunction with a dose reduction or a "hold" of warfarin - even if the procedure is ultimately cancelled.
14	Cancer	Any mention of cancer, with or without specific therapies such as chemotherapy, radiation, etc. DOES NOT INCLUDE a mere history of cancer.
15	Missed Appointments	Any recorded missed appointments - unless due to hospitalization (which is a different variable).
16	Systemic Corticosteroids	Any mention. DOES NOT INCLUDE joint injections, skin creams, etc.
17	Alternative Medications	Any mention - including but not limited to saw palmetto, St. John's Wort, Echinacea, Coenzyme Q10, etc.
18	Antibacterial Antibiotics	Any mention - must be systemic therapy, not local (such as skin creams, etc.)

For all items, one mention is sufficient to mark the item "yes." Mark a "1" if present, or a "0" if absent.

reviewed by all three reviewers to assess inter-rater agreement.

Adverse Events

Ischemic stroke/systemic arterial embolism, VTE and major hemorrhage were the adverse outcomes of interest. We defined major hemorrhage according to the definition of the International Society of Thrombosis and Haemostasis: a fatal event, an event requiring hospitalization with transfusion of at least two units of packed red blood cells, or bleeding involving a critical anatomical site such as the cranium or the retroperitoneum [22]. All patient progress notes were individually reviewed for evidence of adverse events; events were validated directly with the sites by McKesson.

Statistical Analyses

Kappa (κ) statistics were computed to assess inter-rater reliability for the second chart review. To assess significance of effects when comparing categorical variables with the matched design, we used Monte Carlo permutation methods with 10, 000 iterations to compute empirical p-values. Case-control status within each "cluster" of matched observations was randomly permuted 10, 000 times, with a test statistic (e.g., Pearson's chi-square statistic) calculated upon each iteration. This was used as a reference distribution, under the null hypothesis of no association with case status, to compute the empirical p-value. Groups were compared on continuous variables using a generalized linear model to account for correlation between each case and its matched controls. We used conditional logistic regression models to determine the factors that independently predict case status while controlling for patient level covariates (i.e. age, gender, race, co-morbid conditions). Analyses were performed using SAS, version 9.1 (SAS Corporation) and R, version 2.8 (R Foundation).

Results

Baseline Characteristics

There were 109 cases with high dose variability and 218 site-matched controls with low dose variability (Table 2). The mean coefficient of variation (CoV) of the cases was 0.24 and the mean for the controls was 0.02 ($p < 0.001$). The two groups were similar in demographics: most participants were white (89% of cases and 94% of controls) and many were 75 years of age or older (50% of cases and 42% of controls). Forty-five percent of cases were female, compared to 32% of controls ($p = 0.02$). Atrial fibrillation

Table 2 Baseline patient characteristics compared between cases (n = 109) and controls (n = 218)

Demographics	Cases (%)	Controls (%)	P-value
Age 75 or Older	50	42	0.20*
Female	45	32	0.02*
Nonwhite Race	11	6	0.10*
Hypertension	48	47	0.99*
Diabetes	21	21	0.99*
Coronary Artery Disease	39	34	0.53*
Follow up time	10.9 months	11.5 months	< 0.001†
# INR/month	2.3	1.2	< 0.001†
Indication:			0.26*
Atrial Fibrillation	67	58	
VTE	11	13	
Valvular Heart disease	6	5	
Other	16	24	

*Comparison via Monte Carlo simulation

†Comparison via Generalized Estimating Equations (GEE)

was the indication for anticoagulation in 67% and 58% of the cases and controls, respectively. Among the 69 patients with "other" indications for anticoagulation, 27 were anticoagulated for stroke, transient ischemic attack, or cerebrovascular disease; 22 were anticoagulated for congestive heart failure; 13 were anticoagulated for coronary artery disease; 4 were anticoagulated for hypercoagulability; and 3 were anticoagulated for other reasons. Co-morbidities such as hypertension, diabetes, and coronary artery disease were similar between groups.

Validation of Coefficient of Variation as a Measure of Risk

CoV corresponded well with other measures of anticoagulation control and risk for adverse events. The 109 case patients had a mean TTR of 53%, compared to 79% for the 218 control patients ($p < 0.001$). Cases had a higher rate of adverse events. Six case patients experienced major hemorrhage, compared to only 1 control patient ($p < 0.001$). Three case patients had thromboembolic events (2 embolic strokes and 1 pulmonary embolism), compared to 0 control patients ($p = 0.04$).

Predictors of Dose Variability

We assessed predictors of dose variability using chart review. The 3 reviewers achieved a very good rate of inter-rater reliability (three-way $\kappa = 0.76$). In the unadjusted analysis (Table 3), most of the risk factors we examined were associated with case status. Particularly strong associations were seen with amiodarone (12 cases vs. no controls, $p < 0.001$) and a diagnosis of cancer (8 cases vs. 1 control, $p < 0.001$). When present, these variables were highly indicative of high CoV.

After adjustment for covariates (Table 4), variables independently associated with large dose variation included hospital/nursing home stay (OR = 21.3), decreased oral intake (OR = 12.2), use of systemic steroids (OR = 6.1), use of acetaminophen (OR = 4.0), and use of antibiotics (OR = 2.7). Effect size of amiodarone and cancer could not be calculated because there were too few controls with these variables. The presence of these variables precluded model convergence; therefore, these variables were omitted from the model.

Table 3 Proportion of cases and controls with risk factors for extreme dose variability (unadjusted results).

Risk Factors	Cases (n = 109)	Controls (n = 218)	p-value*
Acetaminophen	**33 (30%)**	**30 (14%)**	**< 0.001**
Adherence	**48 (44%)**	**67 (31%)**	**0.01**
Alcohol	9 (8%)	16 (7%)	0.99
Alternative Medication	11 (10%)	10 (5%)	0.08
Amiodarone	**12 (11%)**	**0 (0%)**	**< 0.001**
Antibiotic Use	**47 (43%)**	**44 (20%)**	**< 0.001**
Cancer	**8 (7%)**	**1 (0%)**	**< 0.001**
CHF (Decompensated)	**14 (13%)**	**7 (3%)**	**0.001**
Decreased Oral Intake	**21 (19%)**	**8 (4%)**	**< 0.001**
Diarrhea	**15 (14%)**	**10 (5%)**	**0.003**
Dietary Supplement	8 (7%)	6 (3%)	0.06
Dietary Vitamin K	40 (37%)	73 (34%)	0.61
Hospitalizations/Nursing Home	**47 (43%)**	**12 (6%)**	**< 0.001**
Missed Appointments	12 (11%)	22 (10%)	0.99
Nausea/Vomiting	**10 (9%)**	**5 (2%)**	**0.01**
NSAID Use	**19 (17%)**	**18 (8%)**	**0.02**
Procedures	28 (26%)	41 (19%)	0.21
Systemic Steroids	**12 (11%)**	**9 (4%)**	**0.03**

All variables were obtained by chart review and all are yes/no variables.
Boldface variables are significant at the 0.05 level.

*Via Monte Carlo simulation

Discussion

In this study, we have describe a new measure to identify patients at risk for adverse outcomes of anticoagulation care, have shown that the measure is correlated with INR control and adverse events, and have examined patient-level predictors of being in this high-risk group. The characteristics independently predictive of large weekly variation in warfarin dose were hospitalization/nursing home stay, decreased oral intake, use of systemic steroids, acetaminophen, and antibiotics. In addition, the use of amiodarone and a diagnosis of cancer were almost certainly risk factors for high CoV, though we could not estimate an effect size.

This study suggests that CoV could be an important tool for identifying patients at high risk for poorly controlled anticoagulation therapy and adverse events. Patients identified as high-risk might be referred for case management, adherence training, more intensive follow-up, or indeed reconsideration of whether this particular patient is a good candidate for warfarin. The utility of

Table 4 Multivariate analysis of risk factors for extreme warfarin variability

Chart Review Variables	Odds Ratio (95% CI)	p-value
Acetaminophen	**4.0 (1.33 to 6.30)**	**0.01**
Adherence	2.0 (0.87 to 4.65)	0.10
Alcohol	2.5 (0.65 to 10.00)	0.18
Alternative Medication	2.0 (0.38 to 9.63)	0.44
Amiodarone	*	*
Antibiotic Use	**2.7 (1.11 to 6.33)**	**0.03**
Cancer	*	*
CHF (Decompensated)	2.0 (0.34 to 11.58)	0.44
Decreased Oral Intake	**12.2 (2.25 to 65.68)**	**0.004**
Diarrhea	2.8 (0.51 to 15.67)	0.23
Dietary Supplement	1.0 (0.12 to 7.90)	0.98
Dietary Vitamin K	2.1 (0.86 to 4.92)	0.10
Hospitalizations/Nursing Home	**21.3 (6.21 to 73.14)**	**< 0.001**
Missed Appointments	1.6 (0.51 to 5.15)	0.42
Nausea/Vomiting	4.4 (0.70 to 27.91)	0.11
NSAID Use	1.3 (0.36 to 4.72)	0.69
Procedures	1.4 (0.59 to 3.36)	0.44
Systemic Steroids	**6.1 (1.10 to 34.20)**	**0.04**

Variables are adjusted for all other variables in the table, as well as for age, gender, race, and comorbid conditions (not shown).

* These variables were not estimable in the multivariate model, because too few control patients had these characteristics. Therefore, these variables were omitted from the model.

such an approach for preventing adverse events could be examined in a prospective study. Anticoagulation control (as measured by TTR) could also be used to prospectively identify patients at high risk for adverse events. Our study did not directly compare the ability of these two measures (TTR vs. dose CoV) to identify patients at highest risk for adverse events; this would also be a suitable topic for future study. We suspect that, in many care settings, there is no effort to prospectively identify patients at high risk of adverse events. If the utility of this approach can be established, it may be more widely employed.

An ideal next step to further this research would be to use CoV to identify patients at high risk for poor outcomes in the context of a quasi-experimental design. At some sites of care, patients with extremely high CoV might be referred for case management, adherence training, more intensive follow-up, or indeed reconsideration of whether this particular patient is a good candidate for warfarin. At other sites of care, CoV would be noted, but not acted upon. The outcomes for patients with high CoV (TTR and hopefully clinical outcomes) would be compared, and the effectiveness and cost-effectiveness of the intervention assessed.

Hospitalization had the strongest association with unstable anticoagulation control of any variable in our multivariate analysis. Being hospitalized can contribute to variable dosing for several reasons. When patients are hospitalized, warfarin therapy is often interrupted, and patients may receive parenteral anticoagulation or no anticoagulation at all. Hospitalization also involves large changes in the patient's lifestyle and diet. Returning home, the patient attempts to re-establish usual habits while often restarting warfarin therapy at the previous dose. Unsurprisingly, this combination of circumstances produces out-of-range INR values. Hospitalization is also a general marker of illness severity, which can predict poorer anticoagulation control both before and after hospitalization. Previous studies have also examined the event of a hospitalization as a time-dependent inducer of variable anticoagulation control [23].

Several studies have shown an association with warfarin and acetaminophen [24,25]. Hylek et al. [26] described acetaminophen as an underrecognized source of INR elevation. Her study which included a case-control prospective design assessed patients with high INR values (> 6.0). Acetaminophen was noted as a risk factor that was documented only as case studies in the literature previously. One study examined the prevalence of adverse warfarin-drug combinations in a post-mortem toxicology database. Acetaminophen accounted for more than half of the warfarin drug interactions. In that study, there were more deaths with the combination of acetaminophen and warfarin than with either drug alone [27]. Despite these data, discordant findings showing lack of an association with acetaminophen and warfarin potentiation have been reported [28-30]. The present study reinforces the theory that the use of acetaminophen can contribute to poor anticoagulation control.

Several other studies have described factors associated with anticoagulation control [31-33]. One study, similar to ours, examined factors that contribute to unstable control and found no association with dietary habits or the presence of comorbid conditions. Instead, they found greater instability among patients working full-time, among those with inadequate understanding of oral anticoagulation therapy, and among those with CYP2c9*3 variants [31]. Other studies have examined factors associated with extremely stable control. Witt et al. [32,33] performed 2 studies looking at patients that spent 100% of the time in therapeutic range. Both studies found that older age, lack of co-morbidities and a standard INR target range (i.e. 2-3) were associated with stable control.

There are several strengths to our study. We used a large, nationally representative database of patients receiving warfarin in community-based practice. Our three chart reviewers achieved a very good rate of inter-rater reliability. Finally, this database (ACTION) contains weekly warfarin doses for all patients. These data are usually not available, since warfarin is often prescribed

"use as directed" and so dose changes cannot be reliably abstracted. This is a unique feature of this database, without which we could not have performed this study.

Despite this our study has some limitations. First, this study did not address the question of whether high dose variability is a cause or a consequence of poor anticoagulation control, although we would suspect that it is predominantly a consequence of it. Nevertheless, this study does demonstrate that dose variability is both measurable and related to important clinical outcomes, regardless of its causal relationship with anticoagulation control. As such, it might be used to identify patients at elevated risk for adverse events. Second, this study was limited to risk factors for high CoV that were clearly documented in the clinical notes; however, some risk factors may have been present, but poorly recognized or poorly documented. Our results with regard to risk factors for high CoV should be regarded as exploratory, particularly where a risk factor was shown not to predict high CoV, because an absence of documentation is not conclusive proof that something did not occur. Third, we emphasize that we have only subjected our new scoring system to internal validation, i.e. within the same dataset. A higher level of validation would be attained by demonstrating its utility in a separate dataset. Fourth, the confidence intervals identified in our multivariable analysis of patient-level risk factors for high dose variability are quite large. Therefore, the true magnitude of these effects is not precisely known. A final limitation is that this study evaluates patients with a target INR range of 2-3 and at least 1 month of experience with warfarin; our study results may not apply to patients who are new to warfarin or those with other target ranges.

Conclusions

In this study, we have derived and internally validated a new measure to identify patients at high risk for poor anticoagulation control in clinical practice, namely the coefficient of variation of weekly warfarin doses. This measure identifies patients at high risk for poor anticoagulation control and adverse events. Future studies should explore the use of this measure to identify patients for intervention before they have experienced an adverse event.

List of Abbreviations

VTE: Venous Thromboembolism; INR: International Normalized Ratio; TTR: Percent Time in Therapeutic Range; ACTION: The Anticoagulation Consortium to Improve Outcomes Nationally; CoV: Coefficient of Variation; OR: Odds Ratio.

Acknowledgements

AJR is supported by a career development award (CDA-2-08-017) from the United States Department of Veterans Affairs, Health Services Research and Development Service. The authors thank Elaine Hylek for generously providing access to the database.

Disclaimer

The opinions expressed in this manuscript do not necessarily represent the views or policies of the Department of Veterans Affairs.

Author details

[1]Department of Medicine, Boston University School of Medicine, Boston Medical Center, Boston, MA, USA. [2]Biostatistics Section, Children's Hospital of Boston, Boston, MA, USA. [3]Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, Bedford, MA, USA.

Authors' contributions

LM helped conceive the study idea, performed chart reviews, and drafted the manuscript. ME helped conceive the study idea and performed chart reviews. AO performed statistical analyses and provided statistical supervision. LEH helped collect the data, performed statistical analyses, and managed the data. AJR helped conceive the study idea, performed chart reviews, performed statistical analyses, and provided study supervision. All authors participated in interpretation of interim results, made revisions to the manuscript for important intellectual content, and approved the final manuscript.

Authors' Information

LM and ME were third year internal medicine residents at Boston Medical Center at the time this study was performed. The results of this study were presented at the Society of General Internal Medicine's 33rd annual conference in Minneapolis, MN on April 30, 2010.

Competing interests

Data collection for this study was sponsored by Bristol-Myers Squibb. The sponsor was not involved in the study design, study management, data collection, analysis, writing, revision, or decision to submit for publication. The authors do not have any other conflicts of interest to report.

Received: 12 May 2011 Accepted: 10 October 2011
Published: 10 October 2011

References

1. Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G: **Pharmacology and Management of the Vitamin K Antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition).** *Chest* 2008, **133**:160S-198S.
2. Salem DN, O'Gara PT, Madias C, Pauker SG: **Valvular and structural heart disease: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition).** *Chest* 2008, **133**:593S-629S.
3. Singer DE, Albers GW, Dalen JE, Fang MC, Go AS, Halperin JL, Lip GY, Manning WJ: **Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition).** *Chest* 2008, **133**:546S-592S.
4. Kearon C, Kahn SR, Agnelli G, Goldhaber S, Raskob GE, Comerota AJ: **Antithrombotic therapy for venous thromboembolic disease: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition).** *Chest* 2008, **133**:454S-545S.
5. Wells PS, Holbrook AM, Crowther NR, Hirsh J: **Interactions of warfarin with drugs and food.** *Ann Intern Med* 1994, **121**:676-683.
6. Holbrook AM, Pereira JA, Labiris R, McDonald H, Douketis JD, Crowther M, Wells PS: **Systematic overview of warfarin and its drug and food interactions.** *Arch Intern Med* 2005, **165**:1095-1106.
7. Schulman S, Beyth RJ, Kearon C, Levine MN: **Hemorrhagic complications of anticoagulant and thrombolytic treatment: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition).** *Chest* 2008, **133**:257S-298S.
8. Linkins LA, Choi PT, Douketis JD: **Clinical impact of bleeding in patients taking oral anticoagulant therapy for venous thromboembolism: a meta-analysis.** *Ann Intern Med* 2003, **139**:893-900.
9. Rose AJ, Ozonoff A, Grant RW, Henault LE, Hylek EM: **The Epidemiology of Sub-Therapeutic Anticoagulation in the United States.** *Circ Cardiovasc Qual Outcomes* 2009, **2**:591-597.
10. White HD, Gruber M, Feyzi J, Kaatz S, Tse H, Husted S, Albers GW: **Comparison of outcomes among patients randomized to warfarin therapy according to anticoagulant control: results from SPORTIF III and V.** *Arch Intern Med* 2007, **167**:239-245.

11. Connolly SJ, Pogue J, Eikelboom J, Flaker G, Commerford P, Franzosi MG, Healey JS, Yusuf S: **Benefit of Oral Anticoagulant Over Antiplatelet Therapy in Atrial Fibrillation Depends on the Quality of International Normalized Ratio Control Achieved by Centers and Countries as Measured by Time in Therapeutic Range.** *Circulation* 2008, **118**:2029-2037.
12. Jones M, McEwan P, Morgan CL, Peters JR, Goodfellow J, Currie CJ: **Evaluation of the pattern of treatment, level of anticoagulation control, and outcome of treatment with warfarin in patients with non-valvular atrial fibrillation: a record linkage study in a large British population.** *Heart* 2005, **91**:472-477.
13. Veeger NJ, Piersma-Wichers M, Tijssen JG, Hillege HL, van der Meer J: **Individual time within target range in patients treated with vitamin K antagonists: main determinant of quality of anticoagulation and predictor of clinical outcome. A retrospective study of 2300 consecutive patients with venous thromboembolism.** *Br J Haematol* 2005, **128**:513-519.
14. van Walraven C, Oake N, Wells PS, Forster AJ: **Burden of potentially avoidable anticoagulant-associated hemorrhagic and thromboembolic events in the elderly.** *Chest* 2007, **131**:1508-1515.
15. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briet E: **A method to determine the optimal intensity of oral anticoagulation therapy.** *Thromb Haemost* 1993, **69**:236-239.
16. Fihn SD, McDonell MB, Martin DC, Henikoff J, Vermes D, Kent D, White RH: **Warfarin Optimized Outpatient Follow-up Study Group. Risk factors for complications of chronic anticoagulation. A multicenter study.** *Ann Intern Med* 1993, **118**:511-520.
17. Fihn SD, Callahan CM, Martin DC, McDonell MB, Henikoff JG, White RH: **The risk for and severity of bleeding complications in elderly patients treated with warfarin. The National Consortium of Anticoagulation Clinics.** *Ann Intern Med* 1996, **124**:970-979.
18. Rose AJ, Hylek EM, Ozonoff A, Ash AS, Reisman JI, Berlowitz DR: **Patient Characteristics Associated with Oral Anticoagulation Control: Results of the Veterans Affairs Study to Improve Anticoagulation.** *J Thromb Haemost* 2010, **8**:2182-2191.
19. Rose AJ, Ozonoff A, Henault LE, Hylek EM: **Warfarin for atrial fibrillation in community-based practice.** *J Thromb Haemost* 2008, **6**:1647-1654.
20. Garcia D, Regan S, Crowther M, Hughes RA, Hylek EM: **Warfarin maintenance dosing patterns in clinical practice: implications for safer anticoagulation in the elderly population.** *Chest* 2005, **127**:2049-2056.
21. Garcia D, Regan S, Crowther M, Hylek EM: **The risk of hemorrhage among patients with warfarin-associated coagulopathy.** *J Am Coll Cardiol* 2006, **47**:804-808.
22. Schulman S, Kearon C: **Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients.** *J Thromb Haemost* 2005, **3**:692-694.
23. van Walraven C, Austin PC, Oake N, Wells P, Mamdani M, Forster AJ: **The effect of hospitalization on oral anticoagulation control: a population-based study.** *Thromb Res* 2007, **119**:705-714.
24. Gebauer MG, Nyfort-Hansen K, Henschke PJ, Gallus AS: **Warfarin and acetaminophen interaction.** *Pharmacotherapy* 2003, **23**:109-112.
25. Beckey NP, Stevens GR: **The effect of acetaminophen on the international normalized ration in patients stabilized on warfarin therapy.** *Pharmacotherapy* 2007, **27**:675-683.
26. Hylek EM, Heiman H, Skates SJ, Sheehan MA, Singer DE: **Acetaminophen and Other Risk Factors for Excessive Warfarin Anticoagulation.** *JAMA* 1998, **279**:657-662.
27. Launiainen T, Sajantila A, Rasanen I, Vuori E, Ojanpera I: **Adverse interaction of warfarin and paracetamol: evidence from a post-mortem study.** *Eur J Clin Pharmacol* 2010, **66**:97-103.
28. Gadisseur AP, Van Der Meer FJ, Rosendaal FR: **Sustained intake of paracetamol (acetaminophen) during oral anticoagulant therapy with coumarins does not cause clinically important INR changes: a randomized double-blind clinical trial.** *J Thromb Haemost* 2003, **1**:714-717.
29. Fattinger K, Frisullo R, Masche U, Braunschweig S, Meier PJ, Roos M: **No clinically relevant drug interaction between paracetamol and phenprocoumon based on a pharmacoepidemiological cohort study in medical inpatients.** *Eur J Clin Pharmacol* 2002, **57**:863-867.
30. van den Bemt PM, Geven LM, Kuitert NA, Risselada A, Brouwers JR: **The potential interaction between oral anticoagulants and acetaminophen in everyday practice.** *Pharm World Sci* 2002, **24**:201-204.
31. Palareti G, Legnani C, Guazzaloca G, Lelia V, Cosmi B, Lunghi B, Marchetti G, Poli D, Pengo V, ad hoc Study Group of the Italian Federation of Anticoagulation Clinics: **Risks factors for highly unstable response to oral anticoagulation: a case-control study.** *Br J Haematol* 2005, **129**:72-78.
32. Witt DM, Delate T, Clark NP, Martell C, Tran T, Crowther MA, Garcia DA, Ageno W, Hylek EM: **Warfarin Associated Research Projects and other EnDeavors (WARPED) Consortium. Outcomes and predictors of very stable INR control during chronic anticoagulation therapy.** *Blood* 2009, **114**:952-956.
33. Witt DM, Delate T, Clark NP, Martell C, Tran T, Crowther MA, Garcia DA, Ageno W, Hylek EM, WARPED Consortium: **Twelve-month outcomes and predictors of very stable INR control in prevalent warfarin users.** *J Thromb Haemost* 2010, **8**:744-749.

doi:10.1186/1477-9560-9-14
Cite this article as: Marrast *et al.*: Using highly variable warfarin dosing to identify patients at risk for adverse events. *Thrombosis Journal* 2011 **9**:14.

Patient Education and Counseling 85 (2011) e229–e236

Contents lists available at ScienceDirect

Patient Education and Counseling

journal homepage: www.elsevier.com/locate/pateducou

Information Need

Education level, not health literacy, associated with information needs for patients with cancer

Robin K. Matsuyama [a,*], Maureen Wilson-Genderson [b], Laura Kuhn [b], Drew Moghanaki [c], Hetal Vachhani [d], Michael Paasche-Orlow [e]

[a] Department of Social & Behavioral Health, Virginia Commonwealth University (VCU), Richmond, USA
[b] VCU Department of Social & Behavioral Health, Richmond, USA
[c] VCU Department of Radiation Oncology, Cancer Center, Southside Regional Medical Center, Petersburg, USA
[d] VCU, Massey Cancer Center, USA
[e] Boston University School of Medicine, Boston, USA

ARTICLE INFO

Article history:
Received 26 August 2010
Received in revised form 25 March 2011
Accepted 28 March 2011

Keywords:
Education
Educational attainment
Health literacy
Cancer
Information needs

ABSTRACT

Objective: Cancer patients receiving adjuvant therapy encounter increasingly complex situations and decisions with each new procedure and therapy. To make informed decisions about care, they need to be able to access, process, and understand information. Individuals with limited health literacy may not be able to obtain or understand important information about their cancer and treatment. The rate of low health literacy has been shown to be higher among African Americans than among non-Hispanic Whites. This study examined the associations between race, health literacy, and self-reported needs for information about disease, diagnostic tests, treatments, physical care, and psychosocial resources.
Methods: Measures assessing information needs were administered to 138 newly diagnosed cancer patients. Demographics were assessed by survey and health literacy was assessed with two commonly used measures: the Rapid Estimate Adult Literacy in Medicine (REALM) and the Short Test of Health Literacy in Adults (STOFHLA).
Results: Study findings indicate that educational attainment, rather than health literacy, is a significant predictor of information needs.
Conclusion: Overcoming barriers to information needs may be less dependent on literacy considerations and more dependent on issues that divide across levels of educational attainment.
Practice implications: Oncologists and hospital staff should be attentive to the fact that many patients require additional assistance to meet their information needs.

1. Introduction

For patients with limited health literacy, understanding and communicating cancer treatment options, goals, and preferences is particularly challenging [1]. The Institute of Medicine (IOM) defines health literacy as "the degree to which individuals have the capacity to obtain, process and understand basic health information and services needed to make appropriate health decisions" [2]. Up to 98 million Americans struggle to function within the health care system due to limited health literacy [3], making this a significant problem in the delivery of care to the nation's cancer population.

There is compelling evidence that individuals with limited health literacy have worse outcomes, including lower treatment adherence, more frequent hospitalizations, and higher mortality than those with adequate health literacy [4–10]. The definitions and conceptual models of health literacy continue to evolve from initial writings which emphasized reading and math [2,11] to current views that incorporate a broader range of attributes (e.g., listening, communicating, using information) and is well described in a commentary by Berkman et al. [12]. We do know that limited education, low socioeconomic status, and minority race are risk factors for both limited health literacy and worse cancer outcomes [3,4] and may be important mechanisms of disparities in healthcare outcomes for patients with cancer.

Patients are called upon to be increasingly responsible for self-care at the same time that medical care becomes more complex and technical. Studies with breast cancer patients have shown that expressed needs for information about disease and treatment are very high, particularly at the beginning of treatments [13,14]. Since breast cancer is an illness primarily affecting women, we needed to

* Corresponding author at: Department of Social & Behavioral Health, Virginia Commonwealth University (VCU), Massey Cancer Center, PO Box 980149, Richmond, VA 23298-0149, USA. Tel.: +1 804 828 7996; fax: +1 804 828 5440.
E-mail address: Rmatsuyama@vcu.edu (R.K. Matsuyama).

0738-3991/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.pec.2011.03.022

explore other cancer populations. Research indicates that information seeking is associated with better psychosocial outcomes, such as satisfaction with coping, decisions, and treatment as well as improved physical function [15–17]. Patients with lower income and education, and those who are minorities [6], have been shown to ask fewer questions and have greater difficulty effectively communicating preferences [18,19]. This, in turn, negatively affects shared decision making between patients and providers [1,2]. Race and ethnicity also appear to influence information preferences [20] and perceptions of met/unmet information needs [21]. For example, more African Americans than Caucasians request information about psychosocial support (e.g., support groups, family support) from the NCI Cancer Information Service call center [22,23]. African-Americans also tend to receive less information from and communicate less with physicians [19,24,25].

In addition to well-documented disparities that affect access to cancer therapy and survival [26,27], there are disparities in accessibility to and utilization of pertinent cancer care information [18]. Often the individuals most in need of interventions are those that are the hardest to reach or have the least resources. Overall, ethnic minorities and low-income populations have less access than the general population to relevant information necessary to navigate the health care system and obtain optimal care [28]. In order to develop effective information interventions, it is essential to understand what information patients need.

The purpose of the study was to examine and measure the associations between race, health literacy, and self-reported needs for information about disease, diagnostic tests, treatments, physical care, and psychosocial resources. Based on the literature, we expected that African American participants would score lower on health literacy measures [3] and would have more information needs than non-Hispanic White participants [20,29,30]. We hypothesized that patients with lower health literacy would have greater information needs than those with higher health literacy and that health literacy would mediate the impact of race on information needs.

2. Methods

2.1. Participants

Patients eligible to participate in this observational study were newly diagnosed African American and non-Hispanic White adults with solid tumor cancers Stages II–IV and who would be receiving treatment. Patients with stages 0–I cancer were excluded as they may not undergo further treatment and therefore may have fewer and different information needs compared to patients diagnosed at later stages.

2.2. Recruitment procedures

Potential participants were identified through the Virginia Commonwealth University (VCU) Massey Cancer Center's electronic record system and oncologist referral. Patient enrollment included those treated at both the academic campus as well as a community hospital satellite setting. Study staff contacted potential participants' oncologists and requested permission to discuss the study with individual patients. If the oncologist approved, the patient was contacted in-person during a clinic visit or by letter with an opt-out telephone number. People who did not opt-out, received a follow-up phone call to further explain the study and answer any questions.

During the first meeting with patients interested in the study, the consent form was reviewed and written informed consent was obtained for those who agreed to participate. After consent was obtained, study staff administered the questionnaire and participants were paid $25 upon survey completion. This study was approved by the Virginia Commonwealth University Institutional Review Board.

2.3. Measures

2.3.1. Sociodemographics

The questionnaire included sociodemographics, two validated measures of health literacy, and a validated measure of information needs. The sociodemographic variables assessed included: gender, age, race, marital status, educational attainment, total household income in the previous year, cancer type, cancer stage, and insurance status.

2.3.2. Health literacy

Health Literacy assessment was conducted with two commonly used measures. The first, the Rapid Estimate of Adult Literacy in Medicine (REALM) [31], is a 3-min health word pronunciation test. The REALM consists of a list of 66 words printed on a sheet, beginning with common words and becoming progressively more difficult. Participants are asked to read the list aloud and the research assistant judges if words are pronounced correctly. The REALM gives estimates of four levels of proficiency: 3rd grade and below (0–18), 4th to 6th grade (19–44), 7th to 8th grade (45–60), and high school or above (61–66). The REALM has been highly correlated with the Short Test of Functional Health Literacy in Adults (STOFHLA) (0.80), the Test of Functional Health Literacy in Adults (TOFHLA) (0.84), The Wide Range Achievement Test-Revised (WRAT-R) (0.88), Slosson Oral Reading Test-Revised (SORT-R) (0.96), and Peabody Individual Achievement Test-Revised (PIAT-R) (0.97). The REALM test–retest consistency is high at 0.97 [31].

The second measure, The Short Test of Health Literacy in Adults (STOFHLA) [32–34] is a timed paper and pen test, consisting of 36 reading and four numeracy questions. The test assesses reading ability by providing health-related passages with words missing in a modified Cloze format. Respondents choose the correct answer out of four possible choices for each blank. The test assesses numeracy by providing prescription vials and hospital forms. Respondents calculate numeric answers. The reading section (72 points) and the numeracy section (28 points) are summed for the total STOFHLA score, ranging from 0 to 100. The score then is distributed between three health literacy levels: adequate (67–100), marginal (56–66) and inadequate (0–55). Internal consistency is high (Cronbach's alpha 0.98) and has concurrent validity with the long version of TOFHLA ($r = 0.91$) and the REALM [32]. An additional approach to improve sensitivity of the STOFHLA measure was developed by Wolf et al. [35] which divides the raw scores into seven categories. In this study we used both traditional scoring as well as the Wolf Categories.

2.3.3. Information needs

The Toronto Informational Needs Questionnaire (TINQ) was developed to identify information needs of women with breast cancer and tested with 114 women during adjuvant therapy for breast cancer [13,14]. It was adapted for this study by eliminating the four breast cancer specific questions and retaining the remaining 45 questions related to all cancers. Higher scores represent higher information needs. Internal consistency is good between 0.85 and 0.90 for subscales and $\alpha = 0.94$ for the total questionnaire. Participants are asked to rate the importance of possible areas of informational needs on a five point Likert Scale ranging from 1 (not at all important) to 5 (extremely important). They are asked to report their need for information specifically as of the time of survey administration. There are five subscales for

the TINQ. The disease subscale (score range 8–40) is about the disease, expected disease progression, and prognosis. The subscale for diagnostic tests (score range 7–35) addresses purpose, method, and side effects of tests. The treatment subscale (score range 16–80) is related to reasons for, administration of, and reactions to treatment. The physical subscale (score range 6–30) regards self-care. The psychosocial subscale (score range 8–40) addresses emotional and psychological needs for both patient and family. Five questions were added about tangible information needs issues (e.g., "transportation to the cancer center" and "where to find money to pay medical bills"). These tangible information needs questions were generated from findings of focus groups conducted by the investigators [20,36] and pilot tested with 107 African American and White female cancer patients. The TINQ and the tangible information needs questions are scored separately, with the total TINQ score ranging from 45 to 225 and the tangible score ranging from 5 to 25. With the exception of the REALM and STOFHLA, all measures were administered verbally by research staff.

2.3.4. Statistical analysis

Analyses were conducted using PASW 17.0 [37]. Descriptive statistics and bivariate correlations were examined.

Separate regression equations were created to examine the relationship between race, age, gender, cancer type, and cancer stage for the following dependent variables: each information needs subscale (disease, diagnostic tests, treatment, physical, psychosocial and tangible) and total information needs. As health literacy was an independent variable of specific interest, we employed a multistep modeling approach in which health literacy and education were introduced sequentially to distinguish the potentially different effects of these two factors. As such, the next set of equations added a measure of health literacy to test the relationship between health literacy and each of the information needs dependent variables. After examining the relationship between sociodemographics, health literacy, and information needs, we explored the role of education in these relationships by adding education (dummy coded) to each regression equation. As the REALM and STOFHLA measure different health literacy skills and as the traditional STOFHLA and the Wolf STOFHLA categories emphasize different categorical distributions, all models that include health literacy were evaluated three times (i.e., using the REALM, STOFHLA, or the Wolf STOFHLA Categories) to evaluate stability of our findings across different measures of health literacy [35]. Race (dummy coded), age, gender, cancer type, and cancer stage (dummy coded) were included as covariates in each model described above.

3. Results

3.1. Participant characteristics

Participants (N = 138) included 62 (45%) African American and 76 (55%) non-Hispanic White patients with cancer (Table 1). The mean age was 54.7 (SD 11.8) with 62% female and 38% male. Over half the sample had completed education above high school while 25% had attained a high school diploma or General Education Diploma (GED) and 23% had less than high school. Fifty three percent of participants had income less than $30,000 per year with 19% having less than $10,000 per year. While about half of the sample reported being either unemployed (30%) or on disability (22%), most reported having some type of health insurance (75%). Cancer types included gastrointestinal (24%), breast (23%), lung (26%), lymphoma (18%), and other cancer (9%). Stages of cancer included II (35%), III (38%) and IV (27%).

Thirty five percent of the African American sample reported having less than a high school education and 34% reported making less than $10,000 per year. In comparison, 12% of the White sample reported having less than a high school education and 6% reported making less than $10,000 per year.

Table 1
Sociodemographics.

	African American	Non-Hispanic White	Total
N (%)	62 (44.9)	76 (55.1)	N = 138 (100.0)
Age, mean ± SD	52.1 ± 11.9	56.9 ± 11.3	54.7 ± 11.8
Range	21–74	23–80	21–80
Gender, N (%)			
Female	35 (56.5)	51 (67.1)	86 (62.3)
Male	27 (43.5)	25 (32.9)	49 (37.7)
Married, N (%)	21 (33.9)	52 (68.4)	73 (52.9)
Insurance coverage, N (%)			
Insured	41 (67.2)	62 (81.6)	103 (75.2)
Un- or Under-insured	20 (32.8)	14 (18.4)	34 (24.8)
Missing data (N = 1)			
Employment, N (%)			
Employed FT/PT	10 (16.1)	28 (36.8)	38 (27.5)
On disability	16 (25.8)	14 (18.4)	30 (21.7)
Retired	14 (22.6)	14 (18.4)	28 (20.3)
Unemployed	22 (35.5)	20 (14.5)	42 (30.4)
Education, N (%)			
<H.S. Diploma/GED	22 (35.5)	9 (11.8)	31 (22.5)
H.S. Diploma/GED	23 (37.1)	12 (15.8)	35 (25.4)
>H.S. Diploma/GED	17 (27.4)	55 (72.4)	72 (52.2)
Income, N (%)			
<$10,000/year	21 (34.4)	4 (5.5)	25 (18.7)
$10–29,000/year	27 (44.3)	19 (26.0)	46 (34.3)
$30–49,000	5 (8.2)	7 (9.6)	12 (9.0)
>$50,000/year	8 (13.1)	43 (58.9)	51 (38.1)
Missing data (N = 4)			
Cancer type, N (%)			
Gastrointestinal	16 (25.8)	17 (22.4)	33 (23.9)
Breast	10 (16.1)	22 (28.9)	32 (23.2)
Lung	18 (29.0)	18 (23.7)	36 (26.1)
Other	18 (29.0)	19 (25.0)	37 (26.8)
Cancer stage, N (%)			
Stage 2	17 (27.4)	30 (41.1)	47 (34.8)
Stage 3	26 (41.9)	26 (35.6)	52 (38.5)
Stage 4	19 (30.6)	17 (23.3)	36 (26.7)
Missing data (N = 3)			
Health Literacy Score – REALM			(Range 0–63)
Mean ± SD	49.5 ± 18.8	61.3 (4.4)	56.2 ± 14.0
Range	0–63	37–63	0–63
Missing data (N = 14)			
Health Literacy Score – STOFHLA			(Range 0–100)
Mean ± SD	72.8 ± 30.1	90.8 ± 10.4	82.7 ± 23.3
Range	0–98	0–100	0–100

3.2. Health literacy

Fully 86% of participants were found to have Adequate HL on the STOFHLA, while 62% scored at the high school level on the REALM, and 57% scored in the highest level (Category 7) of the Wolf categories. African Americans were more likely to have limited health literacy than Whites, as measured by the REALM (24% vs. 3%, $p < .01$) and the STOFHLA (23% vs. 1%, $p < .01$). Few participants with high school or higher educational attainment scored 6th grade or below on the REALM (3%) or had inadequate or marginal health literacy on the STOFHLA (1.4%) as shown in Table 2.

3.3. Information needs

Information needs were high in all categories (Table 3), with a total mean of 193.3 (SD = 31.7, range 61–225) and an average of 4.3 (Confidence Interval (CI) 95%) out of a possible 5 on the Likert scale. Participants reported the greatest need for information about disease (average 4.6, CI 95%) and the least need for psychosocial information (average 3.7, CI 95%).

Table 2
Health literacy level by education and race.

	African American			White		
	Below H.S. degree	H.S. degree	>H.S. degree	Below H.S. degree	H.S. degree	>H.S. degree
	N (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)
REALM (*N* = 124)						
≤3rd grade	4 (20)	1 (5)	0 (0)	0 (0)	0 (0)	0 (0)
4–6th grade	4 (20)	2 (10)	2 (14)	1 (12.5)	1 (10)	0 (0)
7–8th grade	10 (50)	12 (60)	2 (14)	1 (12.5)	2 (20)	5 (10)
High School	2 (10)	5 (25)	10 (71)	6 (75)	7 (70)	47 (90)
Total	20 (100)	20 (100)	14 (100)	8 (100)	10 (100)	52 (100)
STOFHLA (*N* = 138)						
Inadequate	9 (41)	5 (21)	0 (0)	1 (11)	0 (0)	0 (0)
Marginal	0 (0)	2 (9)	0 (0)	1 (11)	1 (8)	1 (2)
Adequate	13 (59)	16 (70)	17 (100)	7 (78)	11 (92)	54 (98)
Total	22 (100)	23 (100)	17 (100)	9 (100)	12 (100)	55 (100)

Table 3
Participants' information needs.

Toronto Information Needs Questionnaire (TINQ)	African American	White	Total	*p*-Value
TINQ disease	37.2 ± 4.4	35.7 ± 5.5	36.4 ± 5.1	NS
Range	15–40	20–40	15–40	
TINQ diagnostic tests	31.4 ± 5.3	29.4 ± 6.2	30.3 ± 5.9	<.05
Range	11–35	7–35	7–35	
TINQ treatment	73.2 ± 10.6	70.0 ± 12.4	71.4 ± 11.7	*NS*
Range	20–80	16–80	16–80	
TINQ physical	25.7 ± 5.1	23.9 ± 5.3	24.7 ± 5.3	<.05
Range	7–30	6 30	6–30	
TINQ psychosocial	31.7 ± 7.5	28.2 ± 7.0	29.7 ± 7.4	<.01
Range	8–40	8–40	8–40	
TINQ total	199.4 ± 30.3	188.5 ± 32.2	193.3 ± 31.7	<.05
Range	61–225	62–225	61–225	
Tangible information	20.8 ± 4.1	17.1 ± 4.8	18.8 ± 4.9	<.01
Range	10–25	5–25	5–25	

3.4. Bivariate analyses

In the bivariate analyses for information needs (Table 4), lower education significantly correlated with higher information needs in all TINQ subscales, total TINQ, and tangible information. Similarly, African American race was significantly correlated with higher information needs, with the exception of the Disease subscale. The REALM and STOFHLA were significantly correlated with TINQ Psychosocial subscale, TINQ total, and tangible information needs. The Wolf categories were only significantly correlated with TINQ Psychosocial subscale and tangible information needs.

3.5. Regression models

In the set of regression equations separately examining the influence of adding the health literacy measures as predictors of the various information needs, African American race significantly correlated with greater information needs with the exception of TINQ treatment subscale. Gender, age, cancer type, and cancer stage (sociodemographics) were not significant. The health literacy measures were not related to any of the information needs variables. Table 5 shows these results for the STOFHLA; results for the REALM and the Wolf version of STOFHLA scoring are all consistent with these findings and are not shown.

Addition of education to the prediction models reveals that education is associated with each domain of information needs; specifically, a lower level of educational attainment was associated with higher information needs. For example, in the final model, controlling for sociodemographics, health literacy and adjusting for education, we would predict that someone with ≤H.S. education would have a score of 204.3 on the Info Needs Total as compared to a score of 183.8 for someone with HS or more.

The magnitude for the TINQ differences comparing those with less than a H.S. Diploma versus those with more than a H.S. Diploma/GED were statistically significant (i.e., $p < .5$) for the total TINQ as well as for each subscale as seen in Table 4. We computed the effect sizes for low education (≤H.S. degree) as compared to higher education (>H.S. degree) for information needs and the results were as follows: Information needs Disease subscale, Cohen's d = .60, effect size .29; Information needs Tests

Table 4
Bivariate correlations for the Toronto Informational Needs Questionnaire (TINQ), race, health literacy, and education.

Spearman's rho	TINQ disease	TINQ diagnostic tests	TINQ treatment	TINQ physical	TINQ psychosocial	TINQ total	Tangible
Race	−.148	−.196*	−.180*	−.210*	−.264**	−.232**	−.375**
REALM	−.125	−.112	−.127	−.207	−.304**	−.226*	−.357**
STOFHLA	−.048	−.097	−.034	−.053	−.293**	−.173*	−.379**
Wolf Categories	−.022	−.075	−.047	−.068	−.287**	−.155	−.377**
Education	−.250**	−.207*	−.179*	−.233*	−.347**	−.299**	−.430**

Note: Correlations for age, gender, cancer type, and cancer stage were not significant and are not shown in the table.
* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 5
Linear regression equations for Toronto Information Needs Questionnaire (TINQ).

	Model 1[a]			Model 2[b]			Model 3[c]		
	B(se)	CI 95%		B(se)	CI 95%		B(se)	CI 95%	
		Lower	Upper		Lower	Upper		Lower	Upper
TINQ disease (range 15–40)									
Race	−1.9 (.90)*	−3.7	−.10	−1.7 (.98)	−3.6	.26	−.76 (.97)	−2.7	1.2
Cancer stage	−1.1 (.59)	−2.3	.07	−1.1 (.59)	−2.3	.04	−2.6 (.91)**	−4.3	−.74
STOFHLA	–	–	–	−.01 (.02)	−.05	.03	.02 (.02)	−.03	.06
Education	–	–	–	–	–	–	−3.4 (.99)**	−5.4	−1.4
TINQ tests (range 7–35)									
Race	−2.4 (1.0)*	−4.5	−.38	−2.0 (1.1)	−4.2	.25	−.89 (1.1)	−3.1	1.3
Cancer stage	−1.2 (.68)	−2.5	.20	−1.2 (.69)	−2.6	.11	−3.4 (1.0)*	−5.4	−1.3
STOFHLA	–	–	–	−.03 (.03)	−.08	.02	.00 (.02)	−.05	.05
Education	–	–	–	–	–	–	−3.5 (1.1)**	−5.8	−1.3
TINQ treatment (range 16–80)									
Race	−3.9 (2.1)	−8.0	.20	−3.2 (2.3)	−7.6	1.3	−1.2 (2.2)	−5.6	3.2
Cancer stage	−1.7 (1.4)	−4.4	.98	−1.9 (1.4)	−4.6	.85	−6.4 (2.1)**	−10.6	−2.3
STOFHLA	–	–	–	−.05 (.05)	.15	.06	.00 (.05)	−.10	.10
Education	–	–	–	–	–	–	−6.5 (2.3)**	−11.1	−2.0
TINQ physical (range 6–30)									
Race	−2.0 (.93)*	−3.9	−.20	−1.7 (1.0)	−3.7	.28	−.68 (1.0)	−2.7	1.3
Cancer stage	−.55 (.61)	−1.8	.66	−.62 (.62)	−1.8	.61	−1.1 (1.1)	−3.2	1.0
STOFHLA	–	–	–	−.02 (.02)	−.06	.03	.02 (.02)	−.02	.06
Education	–	–	–	–	–	–	−3.4 (1.1)**	−5.5	−1.2
TINQ psychosocial (range 8–40)									
Race	3.6 (1.3)**	−6.2	−1.1	−2.8 (1.4)*	−5.6	−.02	−1.3 (1.4)	−4.1	1.4
Cancer stage	.39 (.85)	−1.3	2.1	.21 (.85)	−1.5	1.9	−1.2 (1.3)	−3.8	1.4
STOFHLA	–	–	–	−.05 (.03)	−.11	.01	−.01 (.03)	−.07	.05
Education	–	–	–	–	–	–	−4.4 (1.4)**	−7.3	−1.6
TINQ total (range 61–225)									
Race	−13.2 (5.7)*	−24.5	−1.9	−10.1 (6.2)	−22.4	2.3	−3.2 (6.0)	−15.2	8.7
Cancer stage	−4.3 (3.8)	−11.8	3.2	−4.8 (3.8)	−12.3	2.8	−17.0 (5.7)**	−28.2	−5.8
STOFHLA	–	–	–	−.18 (.14)		.10	−.02 (.13)	−.28	.24
Education	–	–	–	–	–	–	−21.9 (6.2)**	−34.1	−9.6
Tangible (range 5–25)									
Race	−3.5 (.81)**	−5.1	−1.9	−2.9 (.87)**	−4.6	−1.2	−1.9 (.85)*	−3.6	−.20
Cancer stage	.41 (.53)	−.65	1.5	.28 (.53)	−.77	1.3	−.55 (.80)	−2.1	1.0
STOFHLA	–	–	–	−.04 (.02)	−.07	.00	−.01 (.02)	−.05	.02
Education	–	–	–	–	–	–	−3.2 (.89)**	−4.9	−1.4

Note: Race coded white/African American, African American as referent; education coded <HS/> = HS, <HS referent; cancer stage coded stage 2/> stage 2, stage 2 referent.
[a] Model 1 includes race and cancer stage (data shown) as well as age, gender, and cancer type (data not shown as these variables were not significant in any model).
[b] Model 2 includes all of Model 1 as well as a health literacy measure. Data for STOFHLA models shown. Results for the REALM and Wolf STOFHLA Categories were consistent with these findings and are not reported.
[c] Model 3 includes all of Model 2 as well as education. Results for the REALM and Wolf STOFHLA Categories were consistent with these findings and are not reported.
* Significant at the .05 level (2-tailed).
** Significant at the .01 level (2-tailed).

subscale, Cohen's $d = .60$, effect size .29; Information needs Treatment subscale, Cohen's $d = .53$, effect size .26; Information needs Physical subscale, Cohen's $d = .58$, effect size .28; Information needs Psychosocial subscale, Cohen's $d = .70$, effect size .33; Total TINQ, Cohen's $d = .69$, effect size .33; Tangible information needs, Cohen's $d = 1.01$, effect size .45. There were no meaningful differences for participants with less than a H.S. diploma compared with those who had achieved a Diploma/GED. With the addition of education to the models, none of the measures of health literacy were associated with any domain of information needs. African American race was significantly associated with greater tangible information needs (Cohen's $d = .75$, effect size .35). Stage II cancer was significantly associated with greater information needs in the TINQ Disease, Tests, and Treatment subscales, and the total TINQ, although the effect size was small (<.12) for each subscale.

4. Discussion

4.1. Discussion

Contrary to our hypothesis, health literacy was not significantly associated with information needs, while educational attainment was significantly associated with information needs. That is, people with a low level of educational attainment reported a higher need for cancer care information than people who had completed a higher level of education. Specifically, people who had not completed secondary education had more information needs than those who attained some college or post graduate education. If a higher level of information needs was found to be associated with both limited health literacy and limited education, there would be relatively strong evidence that simplifying educational materials and providing support for patient education is warranted as a means to fulfill patient's self-assessed information needs. As we did not observe such a finding, alternative conceptualizations need to be considered.

There may be several reasons why we observe discordant findings for education and health literacy. First, although these two constructs are related to each other, they measure different phenomenon. Completed education is not always a predictor of either literacy or health literacy [38,39]. Those with limited formal education may have attained higher skills than a grade level would indicate. In addition, educational systems vary and completion of a particular grade level does not ensure equivalent skills across individuals [38,39]. Finally, the meaning of education level may vary by race and actual education (vs. grade completion) may not

be reflected accurately in standard demographic classifications. For example, after the federal ruling in 1955 to end segregation in schools, the Virginia General Assembly cut off state funds from integrated schools and ultimately closed entire school systems in 1958. Some schools were closed for a year, others (like Prince Edward County) were closed until 1964. Over 13,000 students were without school for at least one semester [40]. While public schools were closed for everyone, education remained accessible for the wealthier classes through state funded tuition vouchers to segregated private schools [41]. It is difficult to ascertain the specific effects this has had on those "lost classes" of the late fifties, but what is certain is that the heaviest burden was left on blacks and lower class whites.

Second, while we used the most common health literacy measurement tools used in the medical literature, the REALM is a word pronunciation test and the STOFHLA is a test of prose literacy and numeracy. Other domains of health literacy not evaluated by these measures (i.e., verbal literacy, information finding skills, and navigation) may be more likely connected to information needs for cancer patients. In addition, the measures do not assess other constructs relevant to information needs that may be linked to educational attainment (e.g., self-efficacy, resource rich environments, social status, or other phenomena related to staying in school). It is not clear how lower educational level leads to higher information needs. Additional phenomena – beyond actual literacy skills, such as differing attitudes or expectations, greater social capital, or more resources may be available to individuals whose education level is greater than H.S. than those with a H.S. education or less.

Although there has been a fair amount of work done on health literacy, there are few studies on the role of health literacy in information needs, particularly in cancer [42,43]. It is important to note that the information needs assessment is subjective, i.e., people with higher educational attainment may believe that they have sufficient knowledge about cancer care. They may actually have the knowledge, having accessed information prior to the interview through the Internet, reading pamphlets, or talking with other people. Alternatively, they may not have the knowledge; a false sense of confidence may lead to their low TINQ scores. This could occur because people can be unaware of their information needs or may not be willing to say they need information. Possibly, people with higher education are generally more confident about being able to access information, even if they have not faced the challenges of their current circumstances.

The highest reported need for information was in the disease domain (average 4.6, CI 95%). This domain includes questions about whether cancer will come back, how cancer acts in the body, ways the disease will affect the patient over time, and what caused the cancer. It makes sense that these are critical issues for newly diagnosed patients as they adjust to their diagnosis. Presumably need for information related to disease will decrease over time, once patients have received and processed this information. Information about tests, treatments, and physical self-care may be more easily obtained and understood because they tend to be concrete and finite. For example, once a patient has received a particular test, there would be little need for additional information unless scheduled for a different one.

Two of the greatest differences in information needs by level of education were discovered on the tangible and psychosocial information subscales. These questions relate to issues outside immediate medical concerns. For example, tangible information questions include needing information about transportation to the cancer center and ways to obtain help paying bills. Because lower education was associated with lesser income it is likely that these types of questions were more relevant to individuals who may need additional financial support than to those with adequate means. The psychosocial questions include items about counseling for the patient and family members and finding support groups. Participants with greater education and income were more likely to have health insurance that cover these services and might not need information provided through the cancer center. It may also be more culturally or socially acceptable to participate in therapy or support groups for individuals with greater means than for those with lesser means. Additionally they may have more resources available than do individuals with limited education and income, thus decreasing the need for this information.

As we had posited, African American participants scored lower on the health literacy measures and had some differences in information needs from White participants. In the first set of regression models, African American race was significantly associated with greater information needs. However, when these data are controlled for health literacy and education the difference between races no longer is significant, with the exception of tangible information needs. Generally African Americans in our patient population tend to have lower socioeconomic status than Whites and frequently reside in the city rather than suburbs. This may account for greater tangible information needs related to paying bills and obtaining transportation to their appointments.

Stage of cancer was not significantly associated with information needs until the final step of the regression model, and the effect size was weak. In this sample, patients with Stage II cancers had greater information needs than those with Stage III or IV cancers. Patients with Stage II cancers have more complex choices for curative treatment than those with Stage IV cancers have for palliative treatment, and presumably would require more information to best understand available options. Treatments for Stage III cancers may at times resemble that for Stage II or Stage IV cancers, which may explain the small differences between groups. With a larger sample, important differences between cancer stage may be identified.

4.2. Strengths and limitations

First, this study collected self-report data that we did not attempt to validate by other means. Self report questionnaires are frequently used in behavioral studies and are considered reliable. In this study, the questionnaires were verbally administered by trained interviewers, which has been shown to reveal more detailed and accurate information than written surveys [44].

Second, the study was conducted at a single cancer center. More than half of the sample had limited income (<$30,000/year) and nearly half had a high school education or less so the results may not be generalizable to other populations. However, many teaching hospitals and academic centers serve a similar patient population. The fact that participants primarily were from an underserved and less educated population makes the findings that educational attainment was a greater predictor of information needs than health literacy particularly relevant.

Third, our sample size does not afford adequate power to discern subgroup analyses, such as cancer type, with confidence. The role of cancer stage also needs to be further explored. The sample size is however adequate to judge the main comparison about health literacy and education. In addition, the analyses remained stable across multiple models and analytic approaches.

The majority of the participants scored adequate on the health literacy measures, raising the question of whether the measures lack the sensitivity to identify a more complete construct of health literacy. Although the measures we used are the ones most commonly used in health literacy research they do not comprehensively cover the broader constructs of health literacy [45,46]. Future work to develop measurement tools that capture other dimensions of health literacy could be informative. Non-Hispanic

White participants were more likely than African American participants to have more than a high school education and to have adequate health literacy however when race was removed from the equation the findings did not change.

This study showed that there was some variation between health literacy scores on the STOFHLA (86% adequate health literacy) and the REALM (62% high school level). These findings are consistent with other studies [6,46–48] and described in a systematic review by Paasche-Orlow et al. [6]. Comparison between studies must take into account different ways that researchers categorize "marginal," "inadequate," and "limited" or "low" health literacy.

Finally, findings reported here were from cross-sectional data. As part of an ongoing longitudinal study, continued observation of this cohort with data collection at Time 2 (4 months) and Time 3 (8 months) will provide the opportunity to see how information needs change over time.

4.3. Conclusions

Study findings indicate that educational attainment, rather than health literacy, is a significant predictor of information needs. Although research has demonstrated the benefits of improving the readability of written materials for individuals with limited literacy, overcoming barriers to information needs may be less dependent on literacy considerations and more dependent on issues that divide across levels of educational attainment. More work is needed to understand the attributes that higher education confers and the relevance of those attributes to cancer care information needs. Documentation of education on intake assessment forms may be useful.

Ultimately, there likely will be many barriers to cancer patient's information needs. Identifying these barriers may help with design of systems to promote patient education and fulfillment of educational needs. The current work shows that health literacy, as typically measured in the medical literature, was not an important barrier for the patients in our cohort. Low educational attainment was the sole factor identified; further research is needed to learn how to translate this observation into specific guidance for clinicians. In the meantime, using "Universal Precautions" to ensure patient's comprehension is the most appropriate approach [7].

4.4. Practice implications

Oncologists and hospital staff should be attentive to the fact that many patients may require additional assistance to meet their information needs. Providers should assume that all patients have ongoing information needs until proven otherwise. Providers cannot rely on patients' requests for information but should evaluate and confirm comprehension [49]. Indeed, the absence of questions should be considered a warning sign! More frequent conversations with a patient can help reinforce health concepts and promote a common understanding of the treatment plan. Providers should assume the burden of communication, checking whether they have been clear rather than whether the patient "understands" [50]. Finally, when available, patient navigators, medical librarians, and volunteers can assist both provider and patient by helping to gather and discuss complex information with patients and their support network.

Acknowledgements

This project was funded by the American Cancer Society MRSGT-08-143-01-CPPB.

Disclosure: The first author is a co-investigator on an NIH funded R01 to develop a new health literacy measure for cancer patients.

I confirm all patient/personal identifiers have been removed or disguised so the patient/person(s) described are not identifiable and cannot be identified through the details of the story.

We thank all the many oncologists, nurses and staff who helped us with recruitment and supported the study, Amy Sullivan, EdD for commenting on an earlier draft of this article and Clara Cline for her research on education and desegregation in Virginia.

References

[1] Davis TC, Williams MV, Marin E, Parker RM, Glass J. Health literacy and cancer communication. CA Cancer J Clin 2002;52:134–49.
[2] Nielsen-Bohlman LT, Panzer AM, Kindig DA. Health literacy: a prescription to end confusion. Washington, DC: National Academies Press; 2004.
[3] Kutner M, Greenerg E, Jin Y, Paulsen C. The health literacy of America's adults: results from the 2003 National Assessment of Adult Literacy (NCES 2006-483). Washington, DC: U.S. Department of Education, National Center for Education Statistics; 2006.
[4] Baker DW, Wolf MS, Feinglass J, Thompson JA, Gazmararian JA, Huang J. Health literacy and mortality among elderly persons. Arch Intern Med 2007;167:1503–9.
[5] Dewalt DA, Pignone MP. The role of literacy in health and health care. Am Fam Physician 2005;72:387–8.
[6] Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med 2005;20:175–84.
[7] Paasche-Orlow MK, Schillinger D, Greene SM, Wagner EH. How health care systems can begin to address the challenge of limited literacy. J Gen Intern Med 2006;21:884–7.
[8] Parker RM, Kindig DA. Beyond the Institute of Medicine health literacy report: are the recommendations being taken seriously? J Gen Intern Med 2006;21:891–2.
[9] Rudd RE, Rosenfeld L, Gall V. Health literacy and arthritis research and practice. Curr Opin Rheumatol 2007;19:97–100.
[10] Sentell TL, Halpin HA. Importance of adult literacy in understanding health disparities. J Gen Intern Med 2006;21:862–6.
[11] Baker DW. The meaning and the measure of health literacy. J Gen Intern Med 2006;21:878–83.
[12] Berkman ND, Davis TC, McCormack L. Health literacy: what is it? J Health Commun 2010;5:9–19.
[13] Galloway S, Graydon J, Harrison D, Evans-Boyden B, Palmer-Wickham S, Burlein-Hall S, et al. Informational needs of women with a recent diagnosis of breast cancer: development and initial testing of a tool. J Adv Nurs 1997;25:1175–83.
[14] Graydon J, Galloway S, Palmer-Wickham S, Harrison D, Rich-van der Bij L, West P, et al. Information needs of women during early treatment for breast cancer. J Adv Nurs 1997;26:59–64.
[15] Fallowfield LJ, Hall A, Maguire P, Baum M, A'Hern RP. Psychological effects of being offered choice of surgery for breast cancer. Br Med J 1994;309:448.
[16] Fallowfield L. Acceptance of adjuvant therapy and quality of life issues. Breast 2005;14:612–6.
[17] Ransom S, Jacobsen PB, Schmidt JE, Andrykowski MA. Relationship of problem-focused coping strategies to changes in quality of life following treatment for early stage breast cancer. J Pain Symptom Manage 2005;30:243–53.
[18] Eggly S, Harper FW, Penner LA, Gleason MJ, Foster T, Albrecht TL. Variation in question asking during cancer clinical interactions: a potential source of disparities in access to information. Patient Educ Couns 2010.
[19] Siminoff LA, Graham GC, Gordon NH. Cancer communication patterns and the influence of patient characteristics: disparities in information-giving and affective behaviors. Patient Educ Couns 2006;62:355–60.
[20] Matsuyama RK, Grange C, Lyckholm LJ, Utsey SO, Smith TJ. Cultural perceptions in cancer care among African-American and Caucasian patients. J Natl Med Assoc 2007;99(10):1113–9.
[21] Ramanadhan S, Viswanath K. Health and the information nonseeker: a profile. Health Commun 2006;20:131–9.
[22] Rutten LJ, Squiers L, Hesse B. Cancer-related information seeking: hints from the 2003 Health Information National Trends Survey (HINTS). J Health Commun 2006;11:147–56.
[23] Squiers L, Finney Rutten LJ, Treiman K, Bright MA, Hesse B. Cancer patients' information needs across the cancer care continuum: evidence from the cancer information service. J Health Commun 2005;10:15–34.
[24] Gordon HS, Street Jr RL, Sharf BF, Kelly PA, Souchek J. Racial differences in trust and lung cancer patients' perceptions of physician communication. J Clin Oncol 2006;24:904–9.
[25] Gordon HS, Street Jr RL, Sharf BF, Souchek J. Racial differences in doctors' information-giving and patients' participation. Cancer 2006;107:1313–20.
[26] Ries LAG, Harkins D, Krapcho M, Mariotto A, Miller BA, Feuer EJ, et al. SEER Cancer Statistics Review, 1975–2003;2007.
[27] Groman R, Ginsburg J. American College of Physicians. Racial and ethnic disparities in health care: a position paper of the American College of Physicians. Ann Intern Med 2004;141:226–32.

[28] Safeer RS, Keenan J. Health literacy: the gap between physicians and patients. Am Fam Physician 2005;72:463–8.
[29] Jean-Pierre P, Fiscella K, Griggs J, Joseph JV, Morrow G, Carroll J, et al. Race/ethnicity-based concerns over understanding cancer diagnosis and treatment plan. J Natl Med Assoc 2010;102:184–9.
[30] Saha S, Freeman M, Toure J, Tippens KM, Weeks C, Ibrahim S. Racial and ethnic disparities in the VA health care system: a systematic review. J Gen Intern Med 2008;23:654–71.
[31] Davis TC, Crouch MA, Long SW, Jackson RH, Bates P, George RB, et al. Rapid assessment of literacy levels of adult primary care patients. Fam Med 1991;23:433–5.
[32] Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999;38:33–42.
[33] Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J Gen Intern Med 1995;10:537–41.
[34] Nurss JR, Parker R, Williams M, Baker D, TOFHL.A.. Test of functional health literacy in adults. Snow Camp, NC: Peppercorn; 2001.
[35] Wolf MS, Feinglass J, Thompson J, Baker DW. In search of 'low health literacy': threshold vs. gradient effect of literacy on health status and mortality. Soc Sci Med 2010;70:1335–41.
[36] Grange C, Matsuyama RK, Ingram K, Lyckholm LJ, Smith TJ. Identifying supportive and unsupportive responses of others: perspectives of African-American and Caucasian cancer patients. J Psychosoc Oncol 2008;26:81–99.
[37] SPSS I. PASW 17.0 (formerly SPSS) 2008.
[38] Berkman ND, Dewalt DA, Pignone MP, Sheridan SL, Lohr KN, Lux L, et al. Literacy and health outcomes 2004; AHRQ Publication No. 04-E007-2.
[39] Kutner M, Greenberg E, Jin Y, Boyle B, Hsu Y, Dunleavy E. Literacy in everyday life results from 2003 National Assessment of Adult Literacy 2007.
[40] Lassiter MD, Lewis AB. The moderates' dilemma: Massive resistance to school desegregation in Virginia. Charlottesville, VA: University of Virginia Press; 1998
[41] Pratt RA. The color of their skin: education and race in Richmond Virginia 1954–89. Charlottesville, VA: University of Virginia Press; 1993.
[42] Garcia SF, Hahn EA, Jacobs EA. Addressing low literacy and health literacy in clinical oncology practice. J Support Oncol 2010;8:64–9.
[43] Amalraj S, Starkweather C, Nguyen C, Naeim A. Health literacy, communication, and treatment decision-making in older cancer patients. Oncology (Williston Park) 2009;23:369–75.
[44] Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: a review and meta-analysis. Am J Public Health 1994;84:1086–93.
[45] Barber MN, Staples M, Osborne RH, Clerehan R, Elder C, Buchbinder R. Up to a quarter of the Australian population may have suboptimal health literacy depending upon the measurement tool: results from a population-based survey. Health Promot Int 2009;24:252–61.
[46] Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Saha S, Snyder A, et al. Variation in estimates of limited health literacy by assessment instruments and non-response bias. J Gen Intern Med 2010;25:675–81.
[47] Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008;23:561–6.
[48] Ginde AA, Weiner SG, Pallin DJ, Camargo Jr CA. Multicenter study of limited health literacy in emergency department patients. Acad Emerg Med 2008;15:577–80.
[49] Volandes AE, Paasche-Orlow MK. Health literacy, health inequality and a just healthcare system. Am J Bioet 2007;7:5–10.
[50] Rudd RE, Anderson JE. The health literacy environment of hospitals and health centers: partners in action: making your healthcare facility literacy-friendly. Cambridge, MA: Harvard University; 2006.

Public Health, Medicine, and Dentistry as Partners in Community Health: A Pioneering Initiative in Interprofessional, Practice-Based Education

Lois McCloskey, DrPH, MPH; Rebecca Condon, MPH; Christopher W. Shanahan, MD, MPH; James Wolff, MD, MPH, MAT; Corinna Culler, DrPH, RDH; Richard Kalish, MD, MPH, MS

Context: As public health challenges grow more complex, the call for professional education to be interprofessional, collaborative, and grounded in real world practice has intensified. **Objective:** In this article, we describe the development, implementation, and results of one pioneering course at Boston University that aims to prepare public health, medical, and dental students for their combined roles in community health settings. **Setting and Participants**: The Schools of Public Health, Medicine, and Dental Medicine jointly offered the course in partnership with 3 community organizations. Participants include MPH, MD, and DMD candidates. **Intervention:** The course design integrates the use of "The Challenge Model" (created by Management Sciences for Health) with training in public health consultation techniques (eg, community-based participatory research, logic models, monitoring and evaluation). Teams of 6 to 8 medical and public health students collaborate with managers and staff of a community health center to address 1 organizational challenge and recommend a sustainability plan. **Results:** Postcourse evaluations revealed that a cross-disciplinary, practice-based education model is feasible and can meet students' learning objectives and exceed expectations of community partners. We overcame formidable obstacles related to the "silo'ed" nature of academic institutions and the competing priorities within overburdened community organizations. We found that sustained project implementation was attained at some but not all sites, yet all sites highly valued the perspective and contribution of student teams. **Conclusion:** Dynamic and replicable, this practice-based education model is adaptable to professional schools whose work intersects in the real world and calls for collaborative leadership.

KEY WORDS: campus-community partnerships, collaborative leadership, community health, competency-based education, dental education, health professions education reform, interprofessional education, medical education, practice-based education, public health education, team-based education

As the public health challenges facing our nation and world grow more complex and inequities more intractable, the call for major reforms in health professional education grows more urgent. In their groundbreaking report, the 2010 *Lancet Commission on the Education of Health Professionals for the 21st Century* points to problems that keep today's curricula for

Author Affiliations: Department of Community Health Sciences (Dr McCloskey and Ms Condon) and Department of International Health (Dr Wolff), School of Public Health; Department of Internal Medicine, Community Health Unit (Dr Shanahan) and Departments of Internal Medicine and Family Medicine (Dr Kalish), School of Medicine; Department of Health Policy and Health Services Research (Dr Culler), Henry M. Goldman School of Dental Medicine, Boston University, Boston, Massachusetts.

The authors thank their community partners for generous support and participation. These include Boston Healthcare for the Homeless Project; Codman Square Community Health Center, Greater Roslindale Medical and Dental Center, South Boston Community Health Center, and South End Community Health Center. In addition, they thank Deans R. Meenan, K. Antman, and J. Hutter from the Boston University Schools of Public Health, Medicine, and Dental Medicine, respectively, for their support of the pioneering initiative, and Leigh Sweet, MD, MPH, for her contribution to the original course design and its promotion. They acknowledge the Maternal and Child Health Bureau Public Health Training Grant (2 T76 MC000017-15) and the Community Health Unit/General Internal Medicine Section/Department of Internal Medicine at the School of Medicine, both of which provided funds to support the participation of organizational partners.

Correspondence: Lois McCloskey, DrPH, MPH, Boston University School of Public Health, 801 Massachusetts Ave, Ste 451, Boston, MA 02118 (loism@bu.edu).

DOI: 10.1097/PHH.0b013e3182165013

J Public Health Management Practice, 2011, 17(4), 298–307

the health professions fragmented and outdated, and graduates, ill-equipped. Chief among the problems are: "the mismatch of competencies to patient and population needs, poor teamwork, narrow technical focus without broader contextual understanding, and weak leadership."[1(1923)] The commission declares a vision and strategy for the education of health care professionals that is multiprofessional, global in outlook, and systems-oriented.[1]

A rich history of education reform precedes the work of the commission. One hundred years ago, the 1910 Flexner Report launched what is deemed the "first generation" of reform, focused on science-based professionalism in medical education.[2] Similar reports followed in nursing[3] and dentistry[4]; and in 1915, the forward-looking Welch-Rose Report argued the need for collaborative public health and medical practice, and created a blueprint for public health and "preventive medicine" education for the century.[5] The "second generation" of reforms, as noted by the *Lancet Commission*, came in the wake of World War II. Government invested heavily in academic health centers, with strengthened research, education, patient care (medicine), and population services (public health). Instructional innovations were focused on problem-based learning and efforts and interdisciplinary integration.[1]

The commission's far-reaching recommendations reflect a culmination of a "third generation" of calls for innovation across all professions—calls for interprofessional and practice-based education,[6-19] academic-community partnerships,[17,20-24] competency-based instruction,[25-35] team-based learning,[36] and leadership development.[9,29,37] Perhaps most important in launching this generation of reform was the 1988 landmark Institute of Medicine Report, *"The Future of Public Health."*[8] As Fineberg and colleagues observed, "the single most influential recommendation of the Institute of Medicine report (was) its insistence that professional education be grounded in "real world" public health."[38(240)] The report spawned myriad funded educational collaborations between medicine and public health, many of them in community settings.[10-12]

However, serious institutional barriers to realizing these lofty reform goals have been well documented[1,15,24,39]—among them, "hyperspecialization and rigid tribalism"[1(1944)] within professions. Nonetheless, the goals are repeatedly declared central to equipping a health workforce to meet the demands of a culturally diverse and aging population, the burden of chronic illness, intransigent health disparities, and increasingly complex health systems.[1,7,9,15,16]

In this article, we describe a pioneering initiative at Boston University that seeks to go beyond traditional, silo'ed, and institution-bound curricula and demonstrate the feasibility and value of instruction that is multi- and interprofessional; practice-, team-, and competency-based, carried out in community settings. "Leading Community Health Initiatives: Public Health, Medicine and Dentistry as Partners" is a 1-semester 4-credit course offered jointly by the Boston University School of Public Health (BUSPH), Boston University School of Medicine (BUSM), and Boston University School of Dental Medicine (BUSDM), co-taught by faculty from each school, and codriven by the needs of community partners and academic learning objectives. It is the first known course of its kind to be reported in the literature. We have offered and evaluated the course twice, first with the BUSPH and BUSM (fall 2009), and then with the addition of BUSDM (fall 2010). We present here our methods (course development, design, and evaluation), results (evaluation findings), lessons learned, and implications for expansion and replication.

Methods

Setting

Boston University Medical campus is home to 3 professional schools (BUSM, BUSDM, and BUSPH) and Boston Medical Center, the largest safety net hospital in New England. Boston Medical Center is also affiliated with 15 community health centers throughout Boston, serving more than a quarter million people annually. All 3 schools have a strong focus on community health practice. Despite these strengths, the course described here is the first multischool offering and the first in each school to prepare students for interprofessional, team-based practice in the context of community partnerships.

Course development

The course was created as a result of student activism. In the spring of 2008, 2 students—1 from medicine and 1 from public health—participated in a weekend-long leadership development workshop for medical and public health students taught by a BUSPH faculty member. Inspired by collaborative educational experience and what they learned about leadership, the 2 students first met with the deans of education at both schools, and then garnered the support of faculty. They encountered both enthusiasm and resistance. Administrators were concerned about where tuition dollars would go, as no infrastructure or precedent existed for sharing tuition dollars in the context of a single course. Faculty members were concerned about workload and teaching credit. Ultimately, 4 faculty members (2 BUSM, 2 BUSPH) signed on because the course fit their

professional goals and commitments. Both sets of administrators saw the pedagogic and long-term institutional value of the initiative and created the necessary administrative mechanisms. The course is the first to be "double coded" with a separate designation for each school. As a result, tuition dollars and faculty teaching credit flow to the students' home school and department.

The course was reviewed by the curriculum committees of BUSM and BUSPH; both enthusiastically approved it and raised one caution—assure ways to advance the projects beyond the semester's end (eg, student internships, plans, or both for sustainability). These administrative, financial, and curricular supports, which entailed out-of-the-ordinary upfront costs to each school, particularly in faculty time), made it possible for us to launch the course and test its feasibility, value, and sustainability.

Course Design and Implementation

Course goals and learning objectives

The course serves 3 broad goals common to all 3 participating fields and schools: (1) to increase and promote collaboration between medicine, public health, and dentistry; (2) to strengthen partnerships with community organizations in nearby neighborhoods; and (3) to prepare health care professionals to be self-reflective and collaborative team members and leaders, able to affect change in health systems and in the health of communities. The key learning objectives of the course address interdisciplinary/crosscutting competencies within the framework of the Association of Schools of Public Health[28] (leadership, diversity and culture, communication, and systems thinking). These align with key overarching principles of the American Association of Medical Colleges[26]; competencies defined by the Accreditation Council of Graduate Medical Education[27] (practice-based learning and improvement, interpersonal and communication skills, and systems-based practice); and competencies of the American Dental Education Association (ADEA) (health promotion and disease prevention, community involvement, professional growth and development; see Table 1).[34]

Teaching approach and techniques

The course is based on a collaborative team model and uses student-centered, practice-based learning methods, with limited didactic teaching. At the heart of our instruction is the question posed to each student: "How can I, as an emerging physician, dentist, or public health care professional, work with my interprofessional team and community partners to address an organizational challenge that will contribute to improvements in community health?" As such, self and team reflection are critical components of learning.

The course is divided into 4 overlapping units: (1) key themes: complements and tensions across the 3 professions; community partnerships; self-reflective team building and leadership; (2) tools for community health project management and consultation (see later); (3) field work at community sites; and (4) presentations, reflections, and wrap up.

Teams are composed of 6 to 8 students and each is partnered with a community site and preceptor. Each faculty member guides the work of one team and serves as liaison to the respective site. With 4 faculty members, the course can accommodate up to 32 students. We assign students to groups on the basis of student choice and interprofessional balance. In its first year, 27 students enrolled, including 22 in public health and 5 in medicine; and in the second year, 32 enrolled, with 23 in public health, 4 in medicine, and 6 in dentistry. Clinicians among the public health students helped balance the public health/clinical presence in each group. The relatively low number of medical students relates not to lack of interest, but schedule constraints—heavy course load in the first year and a 12-week elective schedule in the fourth year that does not perfectly map onto the 15-week semester at BUSPH.

The 5 community partner sites include 4 federally qualified community health centers and 1 community-based organization dedicated to homeless families (see Table 2). The sites were selected based on: (1) faculty's preexisting relationships with the organization and senior staff; (2) capacity of key staff to precept teams (at least 2 hours per week on average) and engage other staff as needed; (3) identification of a health- or systems-related "problem area" that is a priority of the organization and is feasible to address in a 12-week period. Faculty members meet with site preceptors several months before the course begins to an issue and prepare them for their role. Each is paid an honorarium of $ 1500.

Course elements

To achieve its learning objectives, the course combines tools for leadership development, community health project management, and self-reflection, summarized below:

- **The Challenge Model**—a tool created by *Management Sciences for Health*[40,41] to provide a systematic way for teams to experience the direct impact of leadership practices on achievement of results. It guides students through an iterative process of creating a

TABLE 1 ● Course Learning Objectives and Associated ASPH Competencies for MPH[28]

Learning Objectives[a]	Methods	ASPH Competencies
Increase capacity for collaborative leadership	Self- and team reflection Completion of Challenge Model Hands-on project management Collaboration on team project and presentations	Leadership[b] Develop strategies to motivate and collaborate Leadership Use collaborative methods for achieving organizational and community health goals
Increase capacity engage and partner with diverse community organizations and residents	Community-based participatory research framework and interviewing techniques Conduct meetings, interviews, and focus groups Develop and manage project in partnership Plan for project sustainability	Diversity and Culture[c] Use basic concepts and skills involved in community engagement
Increase ability to present results of collaborative projects to diverse stakeholders	Oral presentation to partner site stakeholders, faculty, and peers Consultant report	Communication and Informatics[d] Demonstrate effective written and oral skills for diverse audiences

Abbreviation: ASPH, Association of Schools of Public Health.
[a]Selected course learning objectives designed to meet ASPH crosscutting competencies.
[b]Aligns with ACGME[27] competency domain, "practice-based learning" and "systems based practice", and ADEA[34] competency domain, "professional growth and development."
[c]Aligns with ACGME competency domain, "interpersonal and communication skills" and ADEA domain, "community involvement."
[d] Same as above.

shared vision, defining a measurable result, scanning the environment, analyzing stakeholder interests, identifying opportunities and obstacles, and articulating a challenge to be met. The team selects priority actions and indicators to monitor their completion. The process requires collaboration with each other and the partner site to assure that the measurable result is feasible in a 15-week period and meaningful to the organization.

- **Logic Model and Data Monitoring and Evaluation Plan**[42]—a conceptual framework that illustrates a project's resources (designated as inputs)

TABLE 2 ● Community Site Partners

Codman Square Community Health Center

Codman Square Community Health Center is located in the heart of Dorchester, one of Boston's poorest and most vulnerable neighborhoods, and is noted for its integration of personal health care with community redevelopment, including a neighborhood gym and school. The population of Dorchester is an ethnically diverse mix of African Americans, European Americans, Caribbean Americans, Latinos, and East and Southeast Asian Americans.

Greater Roslindale Medical and Dental Center

Greater Roslindale Medical and Dental Center services span the full range of primary care and ancillary social services allowing patients to access a comprehensive set of services in one location. Its staff is fluent in Spanish, English, Albanian, and Greek and serves the ethnically diverse, middle- and low-income families primarily from the Roslindale, Hyde Park, and West Roxbury neighborhoods of Boston.

Boston Health Care for the Homeless Project

Boston Health Care for the Homeless Project has provided or assured access to high quality health care for all homeless men, women, and children in the greater Boston area for 25 years. In 2010, the Boston health care for the homeless project served more than 11 000 patients. Most of their clients stay in the emergency shelter system, eat in soup kitchens, or visit drop-in centers. The Boston health care for the homeless project also serves unsheltered homeless people as well as those who were formerly homeless and live in transitional or permanent housing.

South Boston Community Health Center

In addition to a full array of medical services and specialty clinics (eg, diabetes, sports medicine, HIV), South Boston Community Health Center offers a food pantry, interpreter services, financial counseling and a cutting edge, assets-based youth development program, the Institute for a Healthier Community. South Boston Community Health Center serves residents of South Boston, a community in which almost 25% of families live in poverty. Although largely white historically, increasing numbers of Latino and African Americans live in South Boston and are served by South Boston Community Health Center.

South End Community Health Center

South End Community Health Center provides a full range of primary care, mental health, dental, eye care, and women, infants and children services to over 15 000 registered patients. South End Community Health Center is recognized for its outstanding preventive care strategies and its cultural and linguistic accessibility and its deep community roots (53% of the staff is hired from the community, with a Board with high representation of community residents). South End Community Health Center is the largest provider of care for the Latino community in the greater Boston area.

and expected results (designated as outputs) in the short and longer term. A Logic Model depicts assumptions about how a project will influence the factors that underlie the challenge being addressed. Elements of a Logic Model are then linked to a data-monitoring plan for process and outcome evaluation.

- **Community-based participatory research framework**—"a collaborative research approach that is designed to ensure structures for participation by communities affected by the issue being studied, representatives of organizations, and researchers in all aspects of the process to improve health and well-being through taking action, including social change."[43] **Community-based participatory research framework** prompts students to grapple with the personal and ethical challenges inherent in working as academics in racially and culturally diverse community settings. We also train students to use specific tools of **community-based participatory research framework**, such as in-depth interviewing; each develops a semi-structured interview format for use in their stakeholder analysis.
- **E-portfolios and peer reviews:** The original tool we used to facilitate individual reflection and peer review within teams, the *Team Learning Assistant*[44]*;* proved to be too laborious. We replaced it with the use of electronic portfolios.[45] Each student creates an individual electronic portfolio to store their responses to assigned and self-generated reflection questions on his or her learning and professional development. In addition, each team member assesses their own and their teammates' strengths and challenges as a team player and collaborative leader.

Assignments, assessment, and team products

Each team has 4 assignments related to the site project: (1) completed "challenge model" and "logic model"; (2) final presentation to stakeholders at the partner site (rehearsed in class); (3) consultation report that summarizes the process, findings, and products, and makes recommendations for implementation and sustainability; and (4) a team electronic portfolio that visually presents the project and contains all team products. Faculty members provide feedback collectively. Students receive team grades for these 4 assignments and individual grades for class participation and reflections. Table 3 portrays each team's challenge and final products by site for both years.

Evaluation Methods

We used 3 methods to conduct a postcourse evaluation to assess satisfaction and impacts of the course: (1) BUSPH online anonymous course evaluation, including quantitative ratings on course elements and instructors, overall course ratings, and open-ended questions about strengths and limitations; (2) a specially designed online survey for medical and dental students; and (3) interviews with 1 or 2 site preceptors/staff members at each site after project completion. The interviews were conducted by phone and lasted for 15 to 20 minutes, in 2009 by a member of the instructional team 6 months after the semester ended, and in 2010, by a research assistant 6 weeks out. The interviews elicited the perspective of community partners on 3 levels: its impact on the organization, the products created, and the processes by which work was accomplished.[17] Finally, we asked respondents to recommend course improvements and rate their willingness to participate in the future. The Boston Medical Center institutional review board reviewed the evaluation plan and granted an exemption.

Evaluation Results

Student perspectives

Overall, the students highly valued the practice- and team-based model of learning and offered excellent suggestions about how to realize its benefits more fully. Each year approximately 60% of the students completed the online course evaluation. In both years, respondents rated the course highly (4.2 of 5.00 in 2009 and 4.4 in 2010) and the large majority (over 90%) said they would "definitely" or "probably" recommend the course to a friend. Among the specific components assessed, "working with other students was valuable" and "the course connected material to other disciplines" were given the highest rating. Elements related to pacing and workload received lower ratings. Responses to open-ended questions further shed light on the strengths and limitations of the course from student perspective. The 3 most valued aspects repeatedly mentioned were: (1) working closely with a community health organization and stakeholders; (2) being able to solve real life problems; and (3) working on a team with students from other schools, even though it was not easy. One 2009 student stated, "The greatest value was the real-life experience of working with an organization and its stakeholders, and also working in a team, as this is likely to be the case at many points in my career." In 2010, a student put it this way: "Finally-an MPH course that lets us use our skills in the field! So important for our future experiences and jobs. . .I have never learned so much in 15 weeks as I did in this course." Another elaborated, "Giving students the opportunity to take a problem and figure out a solution themselves. We rarely get that much decision-making power in school." Other

TABLE 3 ● Products Recommendations by Community Partner Site

Community Partner Site	Team Challenge	Products & Recommendations
Boston	2009	2009
Healthcare for the Homeless	Incomplete acquisition of immunization data from homeless children	Revised patient consent form Improved data collection tool Feedback loop
	2010	2010
	Low smoking cessation rates among adult clients	Team of smoking cessation champions Organizational-wide culture change recommendations Revised patient flow model
Codman Square Health Center (Team 1)	2009	2009
	Low enrollment rates of pregnant women in the Start Initiative, an infant mortality reduction project	Informational brochure and poster Updated refusal survey Recommendations to expand use of incentives, align data systems and tailor enrollment pitch of the Healthy Start Initiative to diverse groups
Codman Square Health Center (Team 2)	2009	2009
	Low knowledge of Human immunodeficiency virus and compliance with treatment among diverse immigrant patients	Patient educational tool Healthy living card
Greater Roslindale Medical and Dental Center	2009	2009
	Increasing rates of obesity in patients of all ages in Roslindale	Template for "Roslindale Right Bites" brochure and restaurant pitch script Recruitment of participating restaurants with window logo and specific "Right Bites" menus
South Boston Community Health Center	2010	2010
	Food insecurity and lack of affordable healthy food options	Youth-designed photo-voice exhibit on food availability and choices Youth-run game table at South Boston "Community Day" Initiation of community cookbook
South End Community Health Center (Team 1)	2010	2010
	Poor pediatric asthma management	"Childhood Asthma Clinic" membership card and brochure Enhanced workflow map Handbook of community resources Recommendations to enhance patient adherence
South End Community Health Center (Team 2)	2010	2010
	High prevalence of early childhood dental caries	Piloted, standardized protocol for applying fluoride varnish to children under the age of 5 years On-site training for pilot week Reference binders with educational materials for parents Referral cards for dental clinic

strengths frequently noted include the following: "the interactions on our team between medical and public health students," "multiple instructors," and "receiving constant feedback from the professors."

The limitations most frequently noted were: transportation difficulties to sites far from campus (2009 only) and the pacing of the course. In both years, even after significant revisions in 2010, students expressed frustration that they did not "get to the field" sooner and recommended condensing the first unit concerning the themes, defining the challenge and selecting a measurable result.

Over the 2 years, half of the medical and dental students responded to the separate evaluation survey. All but one "agreed" or "strongly agreed" that the course better prepared them to work as a community health professional. As one dental student stated, "I gained a deeper understanding of the current relationship between dentistry and public health, and gained skills to improve it as a practicing dentist." Like their public

health peers, clinical students most valued the immersion into an organization's inner-workings. One medical student expressed it this way: "I think the biggest impact on me was the sense of how organization and implementation of change in a health care setting is accomplished." In both years, students noted that a more equal distribution of students across professions would enhance the course.

Site partner' perspectives

In interviews with 12 site partners, all conveyed high satisfaction with the process and results. They consistently commented on students' professionalism and work quality. When asked to rate the value of the team's presentation and report, all respondents assigned a "4" or "5." When asked to name the most valued aspects, the large majority cited the fresh, focused, outside perspective. In 2009, one nurse manager stated, "We had not taken the time to look at our problem from an organizational point of view and we had not analyzed our data." In 2010, one preceptor noted, "their presence was able to catapult the issue higher on the organizational radar." Speaking from her position as director of public health, one respondent took a different angle, "Students brought analytic and critical thinking to the staff. In community health centers staff work hard and it felt like a treat to create space to introspect and think strategically about their work. It was validating and affirming, and brought their work into focus." Our partners also valued information collected from stakeholders never before included, such as patients, restaurant owners, and even providers "down the hall."

When asked about challenges, preceptors repeatedly noted the difficulty of defining a specific project and doing so early in the process. Informed by our experience in 2009, we defined projects more clearly and pitched them at the right level in 2010; yet like students, the preceptors recommended an earlier start on fieldwork. Some also noted the challenges of large team size.

The impact and sustainability of the projects were mixed in 2009 and uniformly positive in 2010. A few examples convey the factors that facilitated higher versus lower impact and sustained effort. In 2009, the challenge selected by the Boston Health Care for the Homeless team was narrowly focused and addressed a "systems issue" (see Table 3). The team created a "feedback loop" that engaged parents in documenting their child's immunizations. The loop has been used successfully for over 1 year. A project at Greater Roslindale Health Center involved student's recruitment of restaurant owners to participate in the "Right Bites" program to encourage healthy menu options. Although deemed by the heath center staff as "beyond expectations," "Right Bites" has not been implemented because of the absence of resources for ongoing work beyond the health center's walls.

In 2010, all projects were considered highly successful and are being sustained (see Table 3). At South Boston Health Center, students were able to launch a new project within an existing youth empowerment organization; and the center has the staff to continue the work. In the preceptor's words, "It sparked a whole new interest in nutrition. When BU came in, it made us realize we need to continue this work. . .and I loved the idea of the kids working with really dynamic (graduate) students. It was an inspiration to them." Likewise, at South End Health Center, one team created a comprehensive plan to address adult obesity and "marketed it" successfully to the staff; and another effectively demonstrated the feasibility of adding oral health prevention activities to primary care visits by bridging the gap between dental and pediatric clinics. As one preceptor said, "Because of (what) they did, the peds department has started using topical fluoride in kids when they go for check ups!"

All interviewees rated their willingness to participate in future years as "very high." Their recommendations for improvements reinforce those of students: (1) carefully define projects at the outset to allow a rapid start; (2) assure a staff person with adequate time and authority to invest in the process; (3) build in a check-in 3 to 6 months postcourse; and (4) involve students in the advancement of projects through ongoing practice. Boston University School of Medicine has since awarded small grant funding to students on 2 teams to support project evaluation and further implementation.

Discussion

The education of community health professionals is in need of reform. Lofty ideals of interprofessional, practice- and team-based, and community-engaged training have received policy attention for decades. In practice, however, the ideals are rarely applied. As the *Lancet Commission* urges, the time is right for innovative initiatives to apply and test these ideals in "real life" within and across universities.[1] The initiative described here is the first known course to apply the principles of reform on one campus and one community. Our modest though pioneering initiative has produced valuable lessons for others who seek imaginative ways to cut across traditional boundaries and educate health professionals for collaborative community practice.

Lesson 1: A multischool, community-partnered, practice-based course is feasible, can meet its learning objectives, and can contribute to change within community health settings

In 2 semesters, our course provided 59 students substantive first-hand experience in the challenges and rewards of working in urban community health. They experienced the highs and lows of life among mission-minded, committed community health professionals, the draw of working on complex, interesting challenges, the constraints of threatened funding cuts and tight resources, the frustrations of underfunded information systems, and the complications of sometimes highly politicized workplaces. They experienced the complexity of understanding and working with diverse collaborators and stakeholders, working across cultural, class, and racial/ethnic differences; and they discovered their own strengths and limitations as team players and leaders. All of these experiences were the focus of student reflections throughout the course.

Lesson 2: Such efforts require administrative support, careful balancing of team- and skill building and field immersion, and generous time investment of faculty and community partners

Within academia, the administrative challenges were important to tackle first. Financial incentives support a silo approach to learning and there exists no template for cross-school education. As described, we overcame these through student and faculty persuasion, and far-sighted administrators who recognized that the course serves a core mission of the 3 schools. The challenge within academia that has proved most difficult to overcome is the recruitment of medical students, largely due to time constraints.

Pedagogically, the toughest challenge is navigating the fine balance between multiple themes and objectives related to crosscutting competencies in the 3 professions (eg, leadership and team building), specific technical skills needed to plan and implement community health projects (eg, logic model building, interviewing), and the time intensity of the field project itself. Our experience suggests that the lions' share of classroom time should be spent on substantive skill building and that field time should be maximized. Competencies such as team building are best woven into field experience, not explored at length in class.

Pragmatically, the rhythm of semester-based education and the "real time" needs of projects geared to systems or behavior change are not easily melded. It is time intensive for site staff and faculty to support fieldwork of students in a telescoped time frame. We learned from interviews, however, that our partners are profoundly committed to the combined agenda of mentoring students as future leaders and getting a job done in their organization. As others have found, the dedication of community partners is a key factor for success in community-campus initiatives.[39]

Lesson 3: An interprofessional, practice-based course can have positive impacts within the partner organizations that go beyond the measurable results and educational objectives

Our evaluation data suggest that the projects most likely to lead to measurable results and sustained effort are: (1) more narrowly focused at the outset; (2) prioritized by the organization; (3) precepted by staff with adequate time to invest and authority to engage other staff; and (4) backed by the will and resources for sustained effort, even if reliant on volunteers and students. Projects with these characteristics are most likely to meet the ultimate aim of practice-based and community-partnered education: reciprocal benefit.[6,20]

We discovered that even in the context of mixed project success, community partners viewed students' contributions as positive and useful. Sites uniformly valued their fresh, outside perspectives that stimulated internal discussions and ideas. Such benefits are often attributed to the best of outside consultants. These contributions are made by students with substantial faculty guidance, and are, in the end, credited to the university.

Lesson 4: An interprofessional course is an effective venue for faculty to model the competencies of leadership we seek to inspire in future health leaders

As faculty, our close communication and collaboration with the site partners allow us to model the day-to-day workings of strong community partnerships. Likewise, the course integrity depends on our ability as a faculty team to work well together as we teach, jointly assess students, and respond to complicated team dynamics. Such team building is challenging for 4 faculty members with intensely busy schedules. Our commitment to weekly meetings or conference calls, and continual email contact, as well as our willingness to self-reflect, has proved critical to the course and team integrity.

Lesson 5: Collaborative, practice-based educational initiatives can be flexible, expandable, and replicable

In "real world" community health settings, various professions, including managers, nurses, social workers, attorneys, occupational and physical therapists all play

important roles. This course model can be expanded by the inclusion of collaborators from any one or more of these professional schools, as our addition of the BUSDM demonstrated. The addition of each profession will greatly enrich the collaborative learning and the array of projects; yet introduce logistic and pedagogic complexities. On the basis of our experience to date, these are complexities well worth taking on.

Conclusion

Emerging leaders of community health must be prepared to meet complex challenges in communities, locally and globally. As such, schools of public health, medicine, dentistry, and others with health-related missions, must assure curricula that give students first-hand immersion in organizations and experience in collaborative practice—across disciplines, sectors, and diverse cultural environments.[7,29] This pioneering initiative at Boston University, the first known of its kind to be implemented and evaluated, is ideally suited as a vehicle for building competencies in crosscutting areas, such as leadership, community engagement and systems thinking, which are shared by public health, medicine, dentistry, nursing, and other health professions.[3,4,6,16,25-28] Dynamic and replicable, the promise of the model lies in its adaptability to any set of professional schools whose work intersects in the real world and calls for collaborative, interprofessional practice, and leadership.

REFERENCES

1. Frenk J, Chen L, Bhutta Z, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet*. 2010;376(9756):1923-1958.
2. Flexner A. *Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching*. New York, NY: The Carnegie Foundation for the Advancement of Teaching; 1910.
3. The Committee for the Study of Nursing Education. *Nursing and Nursing Education in the United States*. New York, NY: The Rockefeller Foundation; 1923.
4. Gies WJ, Pritchett HS. *Dental Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching*. New York, NY: The Carnegie Foundation for the Advancement of Teaching; 1926.
5. Welch WH, Rose W. *Institute of Hygiene: a report to the General Education Board of Rockefeller Foundation.* New York: The Rockefeller Foundation; 1915.
6. Atchison C, Boatright DT, Merrigan D, et al. Demonstrating excellence in practice-based teaching for public health. *J Public Health Manag Pract*. 2006;12:15-21.
7. Roussos ST, Fawcett SB. A review of collaborative partnerships as a strategy for improving community health. *Annu Rev Public Health*. 2000;21:369-402.
8. Institute of Medicine. *The Future of Public Health*. Washington, DC: National Academy Press; 1988.
9. Committee on Assuring the Health of the Public in the 21st Century, Institute of Medicine. *The Future of the Public's Health in the 21st Century*. Washington, DC: National Academies Press; 2002.
10. Lasker RD, Committee on Medicine and Public Health. *Medicine and Public Health: the Power of Collaboration*. New York, NY: The New York Academy of Medicine; 1997.
11. Phillips DF. Medicine-public health collaboration tested. *JAMA*. 2000;283(4):465-467.
12. Agency for Health Care Policy and Research. AHCPR funds project to foster collaboration between medicine and public health. In: *Research Activities Newsletter, No. 202*. Rockville, MD: Agency for Health Care Policy and Research; 1997. AHCPR Publication 97-0024.
13. Hagar M. *Education for More Synergistic Practice of Medicine and Public Health*. New York, NY: Josiah Macy, Jr Foundation; 1999.
14. Berlin L. *Special Tabulation from the American Association of Colleges of Nursing 2001-2002 Database.* Washington, DC: American Association of Colleges of Nursing; 2002.
15. Committee on Educating Public Health Professionals for the 21st Century, Institute of Medicine. *Who Will Keep the Public Healthy? Educating Health Professionals for the 21st Century*. Washington, DC: National Academy Press; 2003.
16. American Association of Medical Colleges. *Taking Charge of the Future: The Strategic Plan for the Association of American Medical Colleges*. Washington, DC: American Association of Medical Colleges; 1995.
17. Community-Campus Partnerships for Health: Transforming Communities & Higher Education. http://depts.washington.edu/ccph/. Accessed October 12, 2010.
18. Axelsson SB, Axelsson R. From territoriality to altruism in interprofessional collaboration and leadership. *J Interprof Care*. 2009;23(4):320-330.
19. Ginsburg L, Tregunno D. New approaches to interprofessional education and collaborative practice: lessons from the organizational change literature. *J Interprof Care*. 2005;19(suppl 1):177-187.
20. Seifer SD. Service-learning: community-campus partnerships for health professions education. *Acad Med*. 1998;73(3):273-277.
21. Driscoll A, Holland B, Gelmon S, et al. An assessment model for service-learning: comprehensive case studies of impact on faculty, students, community and institutions. *Michigan J Comm Svc Learning*, 1996;3:66-71.
22. Haden NK, Hendricson WD, Kassebaum DK, et al. Curriculum change in dental education, 2003-09. *J Dent Educ*. 2010;74(5):539-557.
23. Yoder M. A framework for service-learning in dental education. *J Dent Educ*. 2006;70(2):115-123.
24. Calleson DC, Seifer SD. Institutional collaboration and competition in community-based education. *J Interprof Care*. 2004;18(1):63-74.
25. Council on Linkages: Accomplishments. http://www.phf.org/resourcetools/Documents/COL_accomplishments_.pdf. Accessed May 2, 2011.
26. Association of American Medical Colleges. Scientific Foundations for Future Physicians. Report of the AAMC-

HHMI Committee. Association of American Medical Colleges Web site. https://services.aamc.org/publications/showfile.cfm?file=version132.pdf&prd_id=262&prv_id=321&pdf_id=132. Accessed February 8, 2011.
27. Accreditation Council for Graduate Medical Education. Common Program Requirements: General Competencies. Accreditation Council for Graduate Medical Education Web site http://www.acgme.org/outcome/comp/GeneralCompetenciesStandards21307.pdf. Accessed February 8, 2011.
28. Association of Schools of Public Health Education Committee. *Master's Degree in Public Health Core Competency Model: Final Version 2.3.* Association of Schools of Public Health Web site. http://www.asph.org/document.cfm?page=851. Accessed October 11, 2010.
29. Calhoun JG, Ramiah K, Weist EM, Shortell SM. Development of a core competency model for the master of public health degree. *Am J Public Health*. 2008;98(9):1598-1607.
30. Welton WE, Kantner TA, Katz SM. Developing tomorrow's integrated community health systems: a leadership challenge for public health and primary care. *Milbank Q*. 1997;75(2):261-288.
31. Josiah Macy Jr Foundation. *Revisiting the Medical School Educational Mission at a Time of Expansion*. Charleston, SC: Josiah Macy Jr Foundation; 2008.
32. Cooke, M, Irby, DM, O'Brien, BC, Shulman, LS. *Educating Physicians: A Call for Reform of Medical School and Residency.* Stanford, CA: The Carnegie Foundation for the Advancement of Teaching; 2010.
33. Association of American Medical Colleges. A snapshot of medical student education in the United States and Canada. *Acad Med*. 2010;85:S1-S648.
34. American Dental Education Association. ADEA competencies for entry into the allied dental professions. *J Dent Educ*. 2010;74(7):769-775.
35. DePaola DP. The revitalization of U.S. dental education. *J Dent Educ*. 2008;72(suppl 2):28-42.
36. Haidet P, O'Malley KJ, Richards B. An initial experience with 'team learning' in medical education. *Acad Med*. 2002;77:40-44.
37. Wright K, Rowitz L, Merkle A, et al. Competency development in public health leadership. *Am J Public Health*. 2000;90:1202-1207.
38. Fineberg HV, Green GM, Ware JH, Anderson BL. Changing public health training needs: professional education and the paradigm of public health. *Annu Rev Pub Health.* 1994;15:237-257.
39. Sandy M, Holland BA. Different worlds and common ground: community partner perspectives on campus-community partnerships. *Michigan Journal of Community Service Learning*. 2006;13(1):30-43.
40. Mansour JB, Vriesendorp S, Ellis A. *Managers Who Lead: A Handbook for Improving Health Services*. Cambridge, MA: Management Sciences for Health; 2005.
41. Mansour M, Mansour JB, El Swesy AH. *Scaling up Proven Public Health Interventions Through a Locally Owned and Sustained Leadership Development Programme in Rural Upper Egypt*. Human Resources for health Web site. http://www.human-resources-health.com/content/8/1/1. Accessed October 5, 2010.
42. Program Development and Evaluation. Logic Model. University of Wisconsin-Extension Website. http://www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html, 2005. Accessed January 10, 2011.
43. Viswanathan M, Ammerman A, Eng E, et al. *Community-Based Participatory Research: Assessing the Evidence*. Rockville, MD: Agency for Healthcare Research and Quality; 2004.
44. Deacon CS, Herman ED, Keldsen SZ, Miller JG, Wakefield PA. *The Team Learning Assistant Workbook*. New York: McGraw-Hill; 2005.
45. Digication e-Portfolio. http://digication.com. Accessed January 10, 2011.

Journal of Health Communication, 16:5–8, 2011

ISSN: 1081-0730 print/1087-0415 online
DOI: 10.1080/10810730.2011.605823

Routledge
Taylor & Francis Group

Introduction

Health Literacy Research: Looking Forward

LAUREN A. McCORMACK

Health Communication Program, RTI International, Research Triangle Park, North Carolina, USA

STEVEN R. RUSH

United Health Group Health Literacy Innovations Program, Minnetonka, Minnesota, USA

NAMRATHA R. KANDULA

Division of General Internal Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

MICHAEL K. PAASCHE-ORLOW

Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

We are pleased to present this special issue on Health Literacy of the *Journal of Health Communication*. This marks the second year in a row that the Journal has published a special issue on this topic, demonstrating its ongoing commitment to field of health literacy. The articles in this special issue help to nudge the field forward, illustrating some important advances and future directions, while acknowledging ongoing challenges.

As with the inaugural version, this issue presents findings from the prior year's Health Literacy Annual Research Conference (HARC) which took place in October 2010 in Washington DC. The number of abstract submissions and meeting attendees

This supplement is supported by the following: National Institute on Minority Health and Health Disparities (NIMHD; grant # 2 R13 MD003392-03); Agency for Healthcare Research and Quality (AHRQ); Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD); National Institute of Biomedical Imaging and Bioengineering (NBIB); Section of General Internal Medicine, Boston University Medical Center; OptumHealth; RTI International; and Health Literacy & Learning Program (HeLP), Northwestern University.

Address correspondence to Michael K. Paasche-Orlow, Department of Medicine, Boston University School of Medicine, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA. E-mail: Michael.PaascheOrlow@bmc.org

increased by 20% relative to the first year, reflecting the growing interest of those from diverse disciplines. The goals of HARC include: (1) professional development, (2) advancing the science of health literacy research, and (3) promotion of interdisciplinary research. We aim to create a venue that can help attract young investigators and new ideas and methods to the field. We aim to promote the discourse in this field of inquiry especially regarding core definitional issues, health disparities, and health quality. The varied nature of the research that is relevant to the problem of health literacy makes an interdisciplinary conference incredibly important; bringing such a group of researchers together provides the milieu for new admixtures, new collaborations, and further creativity. A marker of the success we are having in expanding the field towards this goal is that more than half of the presentations at HARC II represented the work of non-MD PhD investigators from a wide array of disciplines. We are hoping for continued growth for HARC III, which will take place October 17–18 in Chicago and will be held in coordination with the International Conference on Communication in Healthcare (ICCH). We hope to see you there (http://www.bumc.bu.edu/healthliteracyconference/)!

In their Commentary, Pleasant et al., describe a research agenda that includes building a new comprehensive approach to the measurement of health literacy and why this is an important task facing health literacy research and practice. Their call for linking health literacy measurement to theory relates to the Commentary by Johnson et al., which is a call to regain a focus on basic research as the basis for effective interventions. It is quite useful to read these Commentaries in concert with a review of the current status of health literacy intervention research. In fact, we include two papers that review health literacy interventions (Sheridan et al. and Allen et al.) as they represent very different approaches to the review and appraisal of the current health literacy intervention literature.

One study underway tests intervention strategies that help educate patients and improve self-care skills in the area of hypertension (Baker et al.). This is one of the few studies to use a randomized control trial study design and look at clinical outcomes. Kandula and colleagues are using an experimental approach to test the effectiveness of teaching strategies to improve patients' recall and retention of information about diabetes management. Other strategies are considered for increasing health literacy including a wellness curriculum for low-income youth (Diamond et al.) and through adult education programs (Freedman et al.).

Several authors share ongoing work in health literacy measurement and methodology (Fransen et al.), including technology-based data collection approaches (Hahn et al.) and how health literacy can affect health care interactions (Manganello et al. and Martin et al.). Rubin and colleagues look at the association between older adults' spoken interactive health literacy and health care experiences among a low-income population. This paper and the paper by Pizur-Barnekow et al. illustrate the practical challenges of measuring interactive health literacy, which is the least studied of all health literacy components. Chin et al. investigate multiple paths to health literacy by exploring the effects of selected cognitive elements in an elderly cohort for two of the most commonly used measures of literacy, while An et al. examine the comprehension of direct to consumer advertising in an elderly cohort.

This special issue examines health literacy as a risk factor for misuse of medication (Shone et al.) and the relationship between health literacy and various intermediate health outcomes including adherence to medications (Osborn et al.). Sentell et al. present the first population-based examination of the prevalence and

associations of health literacy in a Hawaiian sample. Macabasco-O'Connell and Fry-Bowers describe the knowledge and perception of health literacy among nursing professionals. Hardie et al. demonstrate the relationship between health literacy, health services utilization and cost for members of a health plan. This article along with an accompanying editorial by Rush and Paasche-Orlow promote the expansion of health literacy interventions into larger operational settings. A European perspective on how this can happen is presented by Sørensen and Brand.

The papers included in this special issue clearly represent a wide variety of methods and perspectives. Some of the variability represents logical differences that inherently emerge from the research questions; however, it appears that some of the variability represents underlying conceptual disagreements about health literacy. To a certain extent, we, as the editors of this volume, tried not to impose our views. By and large, we tried to allow authors to express themselves. Yet, the process has reinforced our view that the field could greatly benefit from clarity and consensus. Disagreements will likely persist, but they should represent the well examined views of an intellectually curious and vibrant health literacy research community.

References

Allen, K., Zoellner, J., Motley, M., & Estabrooks, P. A. (2011). Understanding the internal and external validity of health literacy interventions: A systematic literature review using the RE-AIM framework. *Journal of Health Communication*, *16*(S3), 55–72.

An, S., & Muturi, N. (2011). Subjective health literacy and older adults' assessment of direct-to-consumer prescription drug ads. *Journal of Health Communication*, *16*(S3), 242–255.

Baker, D. W., DeWalt, D. A., Schillinger, D., Hawk, V., Ruo, B., Bibbins-Domingo, K., et al. (2011). "Teach to Goal": Theory and design principles of an intervention to improve heart failure self-management skills of patients with low health literacy. *Journal of Health Communication*, *16*(S3), 73–88.

Chin, J., Morrow, D. G., Stine-Morrow, E. A. L., Conner-Garcia, T., Graumlich, J. F., & Murray, M. D. (2011). The process-knowledge model of health literacy: Evidence from a componential analysis of two commonly used measures. *Journal of Health Communication*, *16*(S3), 222–241.

Diamond, C., Saintonge, S., August, P., & Azrack, A. (2011). The development of Building Wellness, a youth health literacy program. *Journal of Health Communication*, *16*(S3), 103–118.

Fransen, M. P., Van Schaik, T. M., Twickler, T. B., & Essink-Bot, M. L. (2011). Applicability of internationally available health literacy measures in the Netherlands. *Journal of Health Communication*, *16*(S3), 134–149.

Freedman, A. M., Miner, K. R., Echt, K. V., Parker, R., & Cooper, H. L. F. (2011). Amplifying diffusion of health information in low literate populations through adult education health literacy classes. *Journal of Health Communication*, *16*(S3), 119–133.

Hahn, E. A., Choi, S. W., Griffith, J. W., Yost, K. J., & Baker, D. W. (2011). Health literacy assessment using talking touchscreen technology (Health LiTT): A new item response theory–based measure of health literacy. *Journal of Health Communication*, *16*(S3), 150–162.

Hardie, N. A., Kyanko, K., Busch, S., LoSasso, A. T., & Levin, R. A. (2011). Health literacy and health care spending and utilization in a consumer-driven health plan. *Journal of Health Communication*, *16*(S3), 308–321.

Johnson, S. E., Baur, C., & Meissner, H. I. (2011). Back to basics: Why basic research is needed to create effective health literacy interventions. *Journal of Health Communication*, *16*(S3), 22–29.

Kandula, N. R., Malli, T., Zei, C. P., Larsen, E., & Baker, D. W. (2011). Literacy and the retention of information after a multimedia diabetes education program and teach-back. *Journal of Health Communication*, *16*(S3), 89–102.

Macabasco-O'Connell, A., & Fry-Bowers, E. K. (2011). Knowledge and perceptions of health literacy among nursing professionals. *Journal of Health Communication*, *16*(S3), 295–307.

Manganello, J. A., & Clayman, M. L. (2011). The association of understanding of medical statistics with health information seeking and health provider interaction in a national sample of young adults. *Journal of Health Communication*, *16*(S3), 163–176.

Martin, L. T., Schonlau, M., Haas, A., Derose, K. P., Rosenfeld, L., Buka, S. L., & Rudd, R. (2011). Patient advocacy: Which literacy skills matter most? *Journal of Health Communication*, *16*(S3), 177–190.

Osborn, C. Y., Cavanaugh, K., Wallston, K. A., Kripalani, S., Elasy, T. A., Rothman, R., & White, R. O. (2011). Health literacy explains racial disparities in diabetes medication adherence. *Journal of Health Communication*, *16*(S3), 268–278.

Pizur-Barnekow, K., Darragh, A., & Johnston, M. (2011). "I cried because I didn't know if I could take care of him": Toward a taxonomy of interactive and critical health literacy as portrayed by caregivers of children with special needs. *Journal of Health Communication*, *16*(S3), 205–221.

Pleasant, A., McKinney, J., & Rikard, R. V. (2011). Health literacy measurement: A proposed research agenda. *Journal of Health Communication*, *16*(S3), 11–21.

Rubin, D. L., Parmer, J., Freimuth, V., Kaley, T., & Okundaye, M. (2011). Associations between older adults' spoken interactive health literacy and selected health care and health communication outcomes. *Journal of Health Communication*, *16*(S3), 191–204.

Rush, S. R., & Paasche-Orlow, M. K. (2011). Incorporating health literacy into larger operational environments. *Journal of Health Communication*, *16*(S3), 9–10.

Sentell, T., Baker, K. K., Onaka, A., & Braun, K. (2011). Low health literacy and poor health status in Asian Americans and Pacific Islanders in Hawai'i. *Journal of Health Communication*, *16*(S3), 279–294.

Sheridan, S. L., Halpern, D. J., Viera, A. J., Berkman, N. D., Donahue, K. E., & Crotty, K. (2011). Interventions for individuals with low health literacy: A systematic review. *Journal of Health Communication*, *16*(S3), 30–54.

Shone, L. P., King, J. P., Doane, C., Wilson, K. M., & Wolf, M. S. (2011). Misunderstanding and unintentional misuse of acetaminophen among adolescents and young adults. *Journal of Health Communication*, *16*(S3), 256–267.

Sørensen, K., & Brand, H. (2011). Health literacy—A strategic asset for corporate social responsibility in Europe. *Journal of Health Communication*, *16*(S3), 322–327.

McInnes et al. BMC Research Notes 2011, 4:295
http://www.biomedcentral.com/1756-0500/4/295

BMC Research Notes

SHORT REPORT Open Access

Use of electronic personal health record systems to encourage HIV screening: an exploratory study of patient and provider perspectives

D Keith McInnes[1,2*], Jeffrey L Solomon[1], Barbara G Bokhour[1,2], Steven M Asch[3,4], David Ross[5], Kim M Nazi[6] and Allen L Gifford[1,2,7]

Abstract

Background: When detected, HIV can be effectively treated with antiretroviral therapy. Nevertheless in the U.S. approximately 25% of those who are HIV-infected do not know it. Much remains unknown about how to increase HIV testing rates. New Internet outreach methods have the potential to increase disease awareness and screening among patients, especially as electronic personal health records (PHRs) become more widely available. In the US Department of Veterans' Affairs medical care system, 900,000 veterans have indicated an interest in receiving electronic health-related communications through the PHR. Therefore we sought to evaluate the optimal circumstances and conditions for outreach about HIV screening. In an exploratory, qualitative research study we examined patient and provider perceptions of Internet-based outreach to increase HIV screening among veterans who use the Veterans Health Administration (VHA) health care system.

Findings: We conducted two rounds of focus groups with veterans and healthcare providers at VHA medical centers. The study's first phase elicited general perceptions of an electronic outreach program to increase screening for HIV, diabetes, and high cholesterol. Using phase 1 results, outreach message texts were drafted and then presented to participants in the second phase. Analysis followed modified grounded theory.
Patients and providers indicated that electronic outreach through a PHR would provide useful information and would motivate patients to be screened for HIV. Patients believed that electronic information would be more convenient and understandable than information provided verbally. Patients saw little difference between messages about HIV versus about diabetes and cholesterol. Providers, however, felt patients would disapprove of HIV-related messages due to stigma. Providers expected increased workload from the electronic outreach, and thus suggested adding primary care resources and devising methods to smooth the flow of patients getting screened. When provided a choice between unsecured emails versus PHRs as the delivery mechanism for disease screening messages, both patients and providers preferred PHRs.

Conclusions: There is considerable potential to use PHR systems for electronic outreach and social marketing to communicate to patients about, and increase rates of, disease screening, including for HIV. Planning for direct-to-patient communications through PHRs should include providers and address provider reservations, especially about workload increases.

* Correspondence: keith.mcinnes@va.gov
[1]Center for Health Quality, Outcomes & Economic Research, ENRM VA Medical Center, Bedford, MA, USA
Full list of author information is available at the end of the article

Background

Arguments for expanding HIV screening are compelling. When detected, HIV can be effectively treated with antiretroviral therapy (ART), which improves patient survival, helps prevent HIV transmission, and is cost effective [1-3]. Nevertheless in the U.S. approximately 25% of those who are HIV-infected do not know it [4,5]. Much remains unknown about how to increase HIV testing rates in the US.

Direct outreach to patients via the Internet is a potentially efficient means of educating patients about the importance of HIV screening. Patient electronic personal health record (PHR) systems may be a useful vehicle for such outreach [6-8]. Little is known, however, about how patients and healthcare providers would perceive use of the PHR to disseminate disease screening messages, or whether such messages would increase HIV testing, e.g. by increasing patient knowledge [9], self-efficacy [10], and activation [11]. Additionally this type of outreach could raise patient concerns about privacy of information on the Internet, especially for stigmatized conditions like HIV. Providers may have concerns that workload will increase, or that direct-to-patient outreach circumvents provider authority. While electronic outreach for health purposes is not new, it has largely been evaluated in the context of randomized trials of specific interventions [12-14], or newsletters for which consumers pro-actively register [15]. Little is known about how providers and patients within a large health care organization would perceive large-scale, unsolicited, outreach via an electronic personal health record system to encourage HIV screening.

As the largest provider of HIV care in the U.S., the Department of Veterans Affairs (VA) is well suited for evaluating different methods for increasing HIV testing. The VA already devotes considerable effort to increasing HIV screening rates [16], including clinical reminders in the electronic medical record, provider performance profiling, and reducing paperwork barriers to testing [17-19]. Still, testing rates are sub-optimal, with an estimated 20% to 50% of VA patients with documented risk factors for HIV infection having been tested [20-22]. The VA's electronic PHR, My Health*e*Vet, contains email addresses of nearly 1,000,000 veterans, most of whom (87%) report using VA health care [23]. Thus the VA is an appropriate system for implementing and evaluating large scale electronic outreach for HIV screening. The VA PHR was (and is) evolving rapidly, with new versions released approximately every 6 months (current version is 11.2). In addition, at the time of this study the VA was preparing to adopt new Centers for Disease Control and Prevention (CDC) recommendations for routine, instead of risk-based, HIV testing. Among other things this involved the elimination of the requirement to obtain written patient consent prior to testing. In the context of this rapidly changing landscape we selected methods which would quickly provide VA policy makers with preliminary patient and provider perceptions of the use of the PHR to encourage more HIV screening.

We explored patient and provider attitudes toward an electronic outreach program for HIV screening, based on a PHR platform. We conducted focus groups with patients and providers about HIV testing. We also discussed diabetes and cholesterol screening with participants to assess whether attitudes toward outreach to increase screening depended on the health condition. We explored the acceptability of messages embedded directly in personal emails versus messages posted on the PHR website. Our focus group guides were informed by the Information-Motivation-Behavioral Skills (IMB) model which has guided health promotion and chronic disease management, including HIV [24,25]. We used the model to guide broad categories of questions to include in the focus groups. Qualitative methods were used because, with such new areas of research, it is important to identify salient patient and provider perceptions and themes prior to embarking on larger scale quantitative research [26-28].

Methods

Overview

Four focus groups were conducted between September 2008 and March 2009, in two phases (Figure 1). The first phase explored HIV screening (and other disease screening) in general terms and sought participant suggestions about the content and framing of the electronic outreach messages. Results from this phase guided the investigators in drafting the content of messages. In the second phase, we presented participants the draft text of HIV, diabetes, and cholesterol screening messages in order to explore patient and provider perceptions of realistic content. It was important to compare perceptions of other chronic health conditions to HIV in order to assess whether HIV-related stigma would adversely affect acceptance of HIV screening messages. The message text was based on VA and Centers for Disease Control and Prevention (CDC) screening guidelines, as well as findings from the phase 1 focus groups. The Institutional Review Board of the Edith Nourse Rogers Memorial VA Medical Center, Bedford, MA approved the study including all recruitment methods. Study subjects completed written informed consent prior to participating.

Participants and Setting

A total of 12 patients (6 in each focus group) and 15 providers (6 in one focus group, 9 in the other)

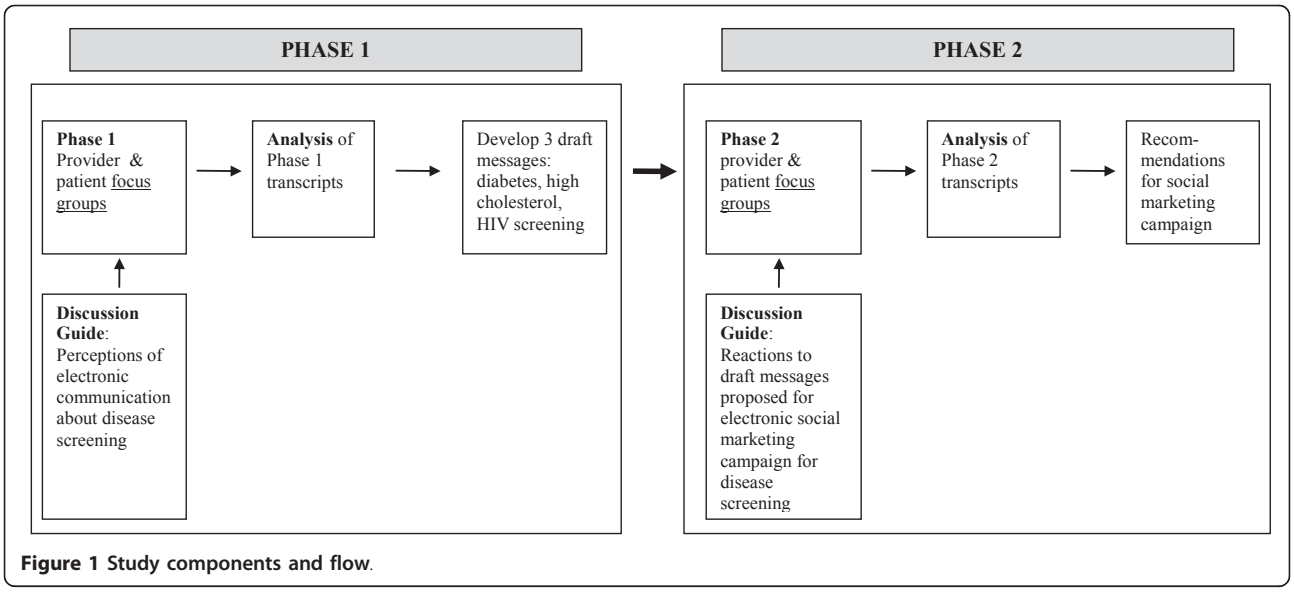


Figure 1 Study components and flow.

participated. Patients were recruited from a Boston area VA medical center. We posted recruitment fliers on walls and a "crawler" message on the televisions in the medical center waiting areas. We also approached veterans at new-patient orientation sessions and in the veterans computer center (KM and LM). Patients received $20 for their participation.

Primary care providers were recruited from another urban New England VA medical center. The invitation to providers was extended by a research team member (KM) who described the study at a primary care staff meeting. Providers were not compensated for participating.

Procedures

There were two facilitators (KM and JS) for each focus group. Following the IMB model, we developed focus group guides to explore whether the concept of electronic disease screening messages, and message content, were perceived as providing valuable information. Secondly questions assessed how likely the information was to motivate patients to seriously consider being tested, and how likely they would be to take action, i.e. ask their provider for a test (behavioral skills). More specifically, the phase 1 patient focus group guide covered experience with disease screening; sources of information about disease screening; experience with Internet and My Health*e*Vet; and, attitudes toward the proposed VA electronic outreach program to increase disease screening rates. The phase 2 patient guide elicited reactions to draft texts of messages for HIV, cholesterol, and diabetes screening (Figure 2) that might be part of the VA disease screening outreach program. Patient focus groups lasted two hours.

The phase 1 provider focus group guide elicited discussion of how providers decide to screen individual patients for diseases; provider views of patient requests for disease screening; provider perceptions of the proposed electronic outreach program; and their assessments of how patients would react to such a program. The phase 2 guide asked providers for their reactions to the same draft messages (for HIV, cholesterol, diabetes) that were shown to patients (Figure 2). They were asked how they anticipated their patients would respond, what questions their patients might ask, and what they thought patient reaction would be. Provider focus groups lasted one hour.

While the research was conducted in two phases, with separate focus group guides for each phase, the results are presented by themes, rather than phase. This is because there was considerable overlap between the focus group guides from the first and second phases. Hence the themes we uncovered emerged from all four focus groups. We have indicated after each quote which phase it came from, e.g. "Patient FG1" refers to first phase patient focus group. Because the focus group guides are lengthy (they make extensive use of probes and prompts that the focus group facilitator can use at his or her discretion) they are not included here, however they are provided for interested readers in Additional File 1: Focus group guides used for patients and providers. All focus groups were audio-taped and transcribed.

Analysis

We used an iterative process to guide the analysis and interpret data, based on grounded theory methods [29]. Immediately following each focus group the facilitators discussed their impressions of significant points that

Dear Veteran:

Did you know that the VA encourages veterans to get a variety of routine health tests, such as checking your cholesterol? Read below to see why.

Why check my cholesterol?

- Over 100 million American adults have cholesterol levels which are higher than recommended.
- Having high blood cholesterol can put you at risk for heart disease, the leading cause of death in the US.
- Adults aged 20 years or older should have their cholesterol checked every 5 years.
- If you think you may not have had a cholesterol test in the past 5 years, ask your provider at your next visit.

The good news!: Cholesterol can be lowered through diet, physical activity, weight control and medication.

Dear Veteran:

Did you know that the VA encourages veterans to get a variety of routine health tests, such as testing for diabetes? Read below to see why.

Why check for diabetes?

- About 24 million Americans have diabetes, but one quarter of these people don't know they have it, because they haven't been tested recently.
- Diabetes can cause serious health problems like heart disease, strokes, blindness, and kidney disease.
- Adults aged 45 years or older should have their blood sugar checked (the test for diabetes) at least every 3 years.
- If you have not had your blood sugar checked in the past 3 years, or if you are unsure, ask your provider at your next visit.

The good news!: If you don't have diabetes, your provider can help you keep it that way. If you do have diabetes, your provider can tell you many ways to control it.

Dear Veteran:
Did you know that the VA encourages veterans to get a variety of routine health tests, such as testing for HIV disease? Read below to see why.

Why check for HIV disease?

- Over 1 million Americans have HIV. Unfortunately a quarter of the people who have HIV don't know they have it because they have never been tested.
- Having HIV but not knowing you have it means that you could spread the virus to other people. Also, untreated HIV causes AIDS, which is a very serious disease.
- The Centers for Disease Control (CDC) recommends that all adults get tested for HIV.
- If you think you may not have been tested for HIV, or are unsure, ask your provider at your next visit.

The good news!: Most people tested for HIV don't have it. But if you do have HIV you won't lose any VA benefits, and the VA has excellent health care for HIV.

Figure 2 Text of electronic messages shown to patients and providers in focus groups.

emerged from the focus group. In addition, within a week the facilitators briefed the whole research team, summarizing the focus group content, and eliciting comments about emerging themes. Audio-recordings were transcribed verbatim by a professional transcription firm. Focus group facilitators (KM and JS) verified the transcripts and analyzed them by open coding, i.e. identifying key concepts emerging from the language used by participants, and assigning codes (descriptive phrases) to segments of text. NVivo

qualitative analysis software (QSR, Melbourne, Australia) was used to facilitate data coding and sorting. Coded text segments were reviewed by three investigators (KM, JS, and BB) to categorize codes into distinct themes. Where similar themes were identified in patient and provider transcripts, we examined similarities and differences in patient and provider perspectives. In a final phase, after developing preliminary interpretations, we searched through the data for alternative interpretations and rival conclusions.

Results

Participant Characteristics

Patients ranged in age from 48 to 71 years of age. Most were white and male. Two-thirds were Internet users (used email and/or the Internet). All had some college education. The 5 physicians and 7 primary-care nurse practitioners participating ranged in age from 46-60 years (see Table 1).

Overview of findings

Patients and providers perceived important informational and educational benefits of the proposed electronic outreach. Several providers expressed substantial privacy concerns related to the social stigma associated with HIV. Patients, for the most part, did not perceive HIV messages to be inherently more sensitive than messages about diabetes and cholesterol. Providers anticipated increased workload and made recommendations for message content in order to minimize disruption to primary care practice.

Perceived benefits for patients of screening messages in general

The more information the better

Patients and providers perceived that electronic disease screening outreach would improve patient access to useful health information, with important educational value. For providers there was a perception that it would reinforce messages they give to patients. Patients seemed interested in more information, and saw this outreach as a potentially good way to achieve this goal. Here, a patient expresses his view that too many people take their bodies and their health for granted, and that the messages proposed could help combat this complacency.

> "I think all this information would be great. Because I think how else are we going to know what to do with the only true asset we own [which] is our body. And some people spend more time getting the oil changed in their car than they do worrying about what's going on in [their bodies]." *(Patient FG2)*

Providers realized that their repeated recommendations to patients to be screened lose effectiveness. Using a new medium, i.e. the Internet, could be a useful adjunct to what providers are trying to communicate to their patients.

> "I think for established patients, this is reinforcing education. The last sentence [of the draft text shown to providers], 'Cholesterol can be lowered,' they're hearing that all the time from us. And now they're reading it, so [it's] another teaching tool." *(Provider FG2)*

Information using lay language and available when patients are ready for it

Patients could imagine scenarios in which disease screening information provided electronically would be better than verbal information from their doctor. The patient below knows there are times when other factors, in this instance substance use, interfere with his ability

Table 1 Characteristics of focus group participants

	Patient Focus Groups	Provider Focus Groups
Number	6 in group 1	6 in group 1
	6 in group 2	9 in group 2*
Gender	2 female; 10 male	8 female; 4 male
Race/Ethnicity	9 white	[Not collected]
	1 African American	
	1 Hispanic	
	1 Pacific Islander/Hawaiian	
Education/Qualifications	6 some college or college degree	7 nurse practitioners
	3 some graduate or graduate degree	5 medical doctors
	3 not provided	
Age	48 to 71 years	46 to 60 years

* There were 12 unique providers because 3 providers participated in both focus groups.

to absorb important messages from his doctor.

> "Let's say I went in from detox. [My doctor] might be saying all this stuff to me but I might be in a situation where I'm like, 'I ain't listening to all this stuff at this point now.' When my head starts to clear out [I might think] 'Okay. What did this doctor say?'" (*Patient FG2*)

An electronic message gives the patient another opportunity to receive the information, and the choice of when and how many times to read it. These messages can be carefully worded to accommodate low literacy levels, as expressed by this patient,

> "If you put [the web information] in layman's terms pretty much explaining LDL or HDL...and how you get it, [that's better than having] $20 dollar words in there." (*Patient FG2)*

Messages can motivate patients

Patients felt that electronic outreach would motivate them to be proactive about their health. Most felt the electronic messages would remind them to be screened, or at least contemplate getting screened. Here a patient finds the idea of an email about HIV screening to be non-threatening, and potentially motivating.

> "They're not telling you [you have to be tested for HIV]. They're putting it in your mind saying... "Have you ever thought about getting HIV testing?" It's non-offensive. You're not prying. But it gets you thinking. Something like that might work." (*Patient FG1*)

Below, two patients, discussing diabetes screening, conclude that outreach messages would be valuable, despite their different perceived risk for the condition. The first realizes that a common "if it's not broken, don't fix it" attitude, may prevent people from thinking about getting preventive testing.

> "As far as [an email] prompting you to go and get [a test] done, yeah there's probably people that aren't even aware that they should have them. Up until five years ago, I never thought about getting my blood sugar checked. What do I care? It's not bothering me any." *(Patient FG2)*

The second patient has a family history of diabetes that he/she might inadvertently ignore. Periodic reminders can be the extra motivation to take action and get tested.

> "My father has diabetes. My mother is borderline diabetes [sic]. I've been checked periodically through the years and I don't seem to be having it...It might slip my mind where I'm not thinking I'm going to get it...and then all of a sudden I see [the electronic message about diabetes screening] and I say, 'Maybe I ought to go and have it checked.' So it's kind of like a kick in the pants." (*Patient FG2*)

HIV content: patient acceptance, provider wariness

Our focus group questions sought to contrast electronic outreach for non-sensitive conditions (i.e. diabetes and cholesterol) with HIV, a stigmatized condition. Few patients, however, made this distinction. Patients thought electronic messages about HIV were acceptable and useful, especially if they were clearly written as public health announcements for wide distribution. One patient likened HIV information delivered electronically through the PHR to posters about HIV testing found in many VA medical center waiting rooms; while another felt that because the material was for a generic patient audience it would not raise objections:

> "I wouldn't mind [getting a message about HIV testing]; it's pasted all over the walls of the VA. I mean, I think the information is good." (*Patient FG2*)
> "None of this is laden with any personal information on yourself or anything like that...I can't see any of this being upsetting to anybody." *(Patient FG2)*

A third patient, however, speaking about messages sent to personal email addresses, was worried about possible security breaches and the stigma of being associated with HIV. He suspected that once information entered his computer it would be difficult to erase, thus allowing later users to find such messages.

> "I don't want 'You get tested for HIV' [in an email]... I've given away computers I've had to people who never had one... They can get into your mainframe, as you folks may know. They can find stuff that you left in there. I'm not taking that chance.... I'm very careful about what goes in my computer. I have a disk that I put everything on. I don't let it go on my mainframe. But some stuff goes in there. You think I want to take a chance and let HIV go in there? And they accidentally find it? Hell, no!" *(Patient FG1)*

This type of concern supports placing the disease screening messages on the PHR website, rather than delivering it directly into patient email inboxes. This sentiment is summarized by a patient in the first focus

group (referring to the PHR by its VA name, "My Health*e*Vet"):

> "I would like to see [a message in my personal email stating] "You have messages at My Health*e*Vet." That's all I want to see. Just tell me to go My Health*e*Vet website, log in and I get messages there. I'd rather see a message there than coming into [my personal email]." *(Patient FG1)*

Providers aired substantially more concerns about HIV messages than patients. Some providers felt that patients would be irresponsible with emails containing HIV-related content. The provider below, for example, described how patients easily find doctor email addresses, and could send their doctors inappropriate email. The provider expresses two issues: the risk that the patient becomes associated, in other people's minds, with a stigmatized condition, and the risk that providers get criticized from their employer for participating in inappropriate email use.

> "I see a lot of problems with this, because there are going to be some [veterans] who aren't thinking about confidentiality. And they're going to be emailing their provider, which is 'my name-dot-VA-dot-gov'; And they're going to be saying 'Oh, I got this thing on HIV. I think I should be tested.' And it's going to be out there in the Internet world, floating around. And the VA is going to get dinged - or me - for 'Oh my God, why did this person email you about this?'" *(Provider FG2)*

This provider expresses the view, correct in some instances, that regular email messages are vulnerable because they "float around" in the Internet easily opened and read by other Internet users.

Another aspect of provider resistance toward HIV-related emails was that they could create suspicions among patients that the VA is withholding information from patients:

> "And I think if you sent them an email, there are some people who might be walking in the next day, 'I got this email that told me to come in and be tested!...Why are you worried?...Why'd you send it to me? Did you send it to anybody else?'" (*Provider FG1*)

This view may reflect provider sensitivity to claims by veterans and active duty military that the US government releases too slowly important health-related information, especially for risks related to military service [30].

Finally, another provider's hesitation was that the HIV message was inappropriate because it was promoting a substantial deviation from the way providers recommended HIV testing. One provider remarked, "...this third [message] on HIV is like a bombshell," because it recommended routine HIV testing. Providers had described in the first focus group that they typically recommend HIV testing to their patients only if risk factors were present, i.e. intravenous drug use or men having sex with men. A consequence of these concerns seemed to be that providers preferred, if an outreach program were conducted, that content be posted on the PHR website, rather than transmitted via email. Patients and providers approved of a "tickler" email message to patients that would indicate there is new content on the PHR, with a hyperlink to the PHR website.

Perceived provider burden

A prominent provider concern was that electronic outreach for disease screening would lead to unmanageable workload. They anticipated the outreach would result in a substantial increase in patient phone calls, time spent explaining and clarifying the outreach program, and additional appointments.

> "If the VA is going to send out a newsletter [about disease screening],...especially if you're sending it electronically,...you're going to get this flood of phone calls the day it goes out, and probably the next week. And, if you're not prepared for that, you've got to have your telephone staff prepared. You have to have your primary care nursing staff prepared, your primary care provider staff. Because these things have this, like, volcano effect." (*Provider FG2*)

One provider suggested that the messages should contain preemptive language to discourage patients from immediately calling or visiting their provider:

> "If you maybe send out [an electronic] newsletter [to patients that says] '... your provider will be asking you for A, B, C, D, E, F, G at your annual - highlighted, underlined, in bold, different color - visit', so [the patients realize] you don't need a PSA every time you come to the walk-in." (*Provider FG1*)

These providers did not reject the electronic outreach initiative, but have suggested that to be successful, it would be wise to make advance preparations with staff and to include education of patients that indicates this is not urgent and can be handled at annual - or other regularly scheduled visits. Other providers concurred, but also reflected a feeling that PCPs are being

shouldered with increasing demands and performance measures, often without increases in resources:

> "How can I do this? I want to be doing X, Y, and Z, and you're adding another element that I'm responsible for." *(Provider FG1)*

Patients, interestingly, did not indicate they would rush to contact their providers or make appointments to see their doctors as a result of electronic messages. In fact some patients believed that if the electronic communication had links to more information it might actually save doctors time:

> "If you need more information...instead of having an hour conversation with the doctor and having the doctor teach you, you could actually go to a place on [the patient website for more information]." (*Patient FG2*)

Discussion

As health care systems adopt new information technologies it is appropriate to consider their use for public health purposes, such as disease screening. This study takes a first, exploratory step in evaluating the acceptability of outreach via an electronic PHR system by soliciting patient and provider perceptions through focus groups. We found that perceptions were, on the whole, positive. Patients and providers acknowledged educational, informational, and motivational benefits of electronic messages. Providers especially, and patients to a much smaller degree, expressed privacy concerns about messages that contained HIV content. Those concerns could be mitigated by posting patient content on the PHR website, as opposed to embedding it in personal email messages. A bigger issue for providers, however, was that this kind of outreach could lead to unacceptable increases in workload. They suggested it could be mitigated by increased primary care resources and management of patient flow so that most additional disease screening could occur during annual visits or spread more evenly over time.

Patients indicated that electronic content afforded the ability to view information when and where convenient, at appropriate reading-levels, and with web-links to multiple sources of information. Using individual email addresses, however, carries the potential risk of creating suspicions among patients that they have been contacted based on specific clinical signs of HIV, or based on HIV risk stereotypes, e.g. homosexuality or intravenous drug use. HIV-related stigma also seemed to underlie provider worries that patients would unwittingly expose themselves to stigma if they sent their doctors emails about HIV testing. Accordingly, patients and providers favored an outreach approach that delivered content impersonally, i.e. posted on the healthcare system PHR website.

Our findings support the IMB model in that both patients and providers indicated that the electronic messages were perceived as providing important information, and that they would lead to patient action in terms of inquiries about, or actual increases in, testing. The findings also highlighted to us that the health belief model (HBM) [31] could be an important addition to the IMB model in helping to understand patient and provider responses to electronic messages about disease screening. This makes sense in that the health belief model often guides health-related social marketing campaigns [32-35] that rely on perceived susceptibility to disease to motivate people to take action. Patients with high perceived susceptibility may seek information, screening, and care on their own. Others patients, however, may have consciously or unconsciously suppressed the knowledge that their family history or risky health behaviors could make them vulnerable to certain health conditions. For such patients the electronic messages serve as external cues ("cues to action" in the terminology of HBM) motivating them to take action and get tested. While the current draft messages (Figure 2) incorporate concepts of information and motivation from the IMB model, future versions could have links to skill-building material - another important IMB component. For example the HIV message could link to material about how to have a conversation with a partner about using condoms, while the diabetes message could link to simple instructions for increasing daily physical activity.

It is noteworthy that the participating patients were middle aged and older adults, most of whom were not highly experienced computer and Internet users. Nevertheless nearly all recognized advantages that such technologies provide in distributing beneficial health information, a finding supported by Pew Research Center findings that older adults are increasing their presence online [36,37].

Provider concerns that electronic communications with patients may create unmanageable workloads have been documented previously [38]. Evidence suggests, however, that patient use of PHRs, secure messaging, and similar electronic communication tools do not overwhelm providers [39,40]. There is even evidence that electronic communication reduces in-person and telephone communication [6]. We found support for this, e.g. a patient stating that accessing information from a website could replace "an hour conversation with the doctor". The above notwithstanding, we do not dismiss provider concerns about increased workloads. Primary

care providers face health system demands for better quality of care at lower costs, with resultant increased stress and loss of autonomy [41,42]. On the other hand solutions exist to even out demands on providers, for example by staggering electronic outreach messages based on patient birth dates or social security numbers.

The success of HIV screening campaigns may rest partly on patient perceptions that, in the event of a positive HIV test, they can gain access to compassionate providers and effective treatments, i.e. there is good linkage to care [43-45]. When HIV screening outreach is conducted by a health facility or system that has strong HIV care programs, it is likely the outreach will be more successful. In this regard, the VA would seem to be especially well suited to employ the kind of electronic outreach described in our study because it is a large, comprehensive health care system with specialized HIV clinics to care for veterans with HIV/AIDS. Currently approximately 23,000 veterans with HIV are in treatment in the VA [46].

This study was conducted 6 months before the VA formally adopted CDC guidelines which recommend routine HIV testing for all adults in care, regardless of risk factors [47]. This policy change eliminated written patient informed consent for HIV testing, in favor of verbal consent. Thus our findings represent patient and provider perceptions prior to implementation of the new HIV testing policy in the VA. Adherence to the CDC guidelines is far from universal even after the policy change in the VA and in other settings [48], suggesting the importance of continued outreach to patients to encourage HIV testing. In addition even in settings where the guidelines are closely adhered too, there will be patients who come for care infrequently and thus would benefit from this kind of outreach; the outreach messages might prompt them to make a visit or a phone call to discuss testing with their provider. In any case it will be important to evaluate whether patient and provider perceptions of HIV testing messages have changed in the VA and to extend the analysis to non VA sites.

As PHR systems continue to expand their capabilities, it is easy to imagine moving from occasional broad electronic outreach programs to more routine patient reminders that patients see when they open up their PHR. In the VA, for example, the PHR has recently implemented reminders for preventive care and procedures, such as diabetes care (foot and retinal exams), cancer screening, and immunizations [49]. It would not be difficult, technically, to add HIV screening to that list.

Limitations

Our study was limited to 2 patient focus groups and 2 provider focus groups conducted in one region of the U. S. Thus our findings may not be generalizable to other regions and other healthcare systems. Only 2 female patients participated (1 in each group), also limiting generalizability. Our participants were middle-aged and older, and thus probably reacted differently to some issues than would participants in their 20s and 30s who have grown up with computers. Also participant responses might have differed had they been reacting to a "live" electronic outreach program rather than a proposed one. Our use of draft text, however, which participants reviewed in the focus groups, is likely to have created a sense of concreteness and immediacy.

Conclusions

The growth in online information systems connecting healthcare organizations with their patients provides an excellent opportunity to conduct low cost and potentially high impact electronic outreach and social marketing. Our findings suggest that patients and providers endorse the use of PHRs for disease screening outreach, even for a stigmatized health condition such as HIV. For providers it is important that prior to initiating wide-scale electronic outreach forethought be given to management of patient expectations and flow. Before large scale implementation of such a program, validation from other geographic regions and with other age groups would be beneficial. If executed properly, electronic outreach campaigns through PHR systems may lead to increased screening, increased detection, and improved health.

Additional material

Additional file 1: Focus group guides used for patients and providers. A text file with the two patient focus group guides and the two provider focus group guides.

Acknowledgements and Funding

This study was supported by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Developoment Service Quality Enhancement Research Initiative (RRP 08-251). Additionally DK McInnes was supported by a VA Career Development Award (CDA 09-016). The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. We wish to thank Linda McCoy, Joanne Anctil, and Jane Burgess for their help with this study. We also thank the patients and primary care providers who took part in the study.

Author details

[1]Center for Health Quality, Outcomes & Economic Research, ENRM VA Medical Center, Bedford, MA, USA. [2]Boston University School of Public Health, Boston, MA, USA. [3]VA Palo Alto Health Care System, Palo Alto, CA, USA. [4]Division of General Medical Disciplines, Stanford School of Medicine, Stanford, CA, USA. [5]VA Public Health Strategic Health Care Group, Washington, DC, USA. [6]VA Veterans and Consumers Health Informatics Office, Washington, DC, USA. [7]Boston University School of Medicine, Boston, MA, USA.

Authors' contributions
KM and AG wrote the initial study protocol. KM and JS conducted the focus groups. JS, BB, and KM anaylzed the data. All authors contributed to interpretation of data. KM wrote the manuscript, which was commented on by all the other authors. All authors have read and approved the final version of the manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 24 June 2011 Accepted: 15 August 2011
Published: 15 August 2011

References

1. Palella FJ Jr, Deloria-Knoll M, Chmiel JS, Moorman AC, Wood KC, Greenberg AE, Holmberg SD: **Survival benefit of initiating antiretroviral therapy in HIV-infected persons in different CD4+ cell strata.** *Ann Intern Med* 2003, **138(8)**:620-626.
2. Centers for Disease Control and Prevention: **Late HIV testing - 34 states, 1996-2005.** *MMWR Morb Mortal Wkly Rep* 2009, **58(24)**:661-665.
3. Kitahata MM, Gange SJ, Abraham AG, Merriman B, Saag MS, Justice AC, Hogg RS, Deeks SG, Eron JJ, Brooks JT, *et al*: **Effect of early versus deferred antiretroviral therapy for HIV on survival.** *N Engl J Med* 2009, **360(18)**:1815-1826.
4. Anonymous: **Screening for HIV: recommendation statement.** *Ann Intern Med* 2005, **143(1)**:32-37.
5. Glynn M, Rhodes P: **Estimated HIV prevalence in the United States at the end of 2003. Abstract: T1-B1101.** *Proceedings of the National HIV Prevention Conference* 2005.
6. Chen C, Garrido T, Chock D, Okawa G, Liang L: **The Kaiser Permanente Electronic Health Record: transforming and streamlining modalities of care.** *Health Aff (Millwood)* 2009, **28(2)**:323-333.
7. Silvestre AL, Sue VM, Allen JY: **If you build it, will they come? The Kaiser Permanente model of online health care.** *Health Aff (Millwood)* 2009, **28(2)**:334-344.
8. Kaelber D, Pan EC: **The value of personal health record (PHR) systems.** *AMIA Annu Symp Proc* 2008, 343-347.
9. Amico KR, Barta W, Konkle-Parker DJ, Fisher JD, Cornman DH, Shuper PA, Fisher WA: **The Information-Motivation-Behavioral Skills Model of ART Adherence in a Deep South HIV+ Clinic Sample.** *AIDS Behav* 2009, **13(1)**:66-75.
10. Kalichman SC, Cain D, Fuhrel A, Eaton L, Di Fonzo K, Ertl T: **Assessing medication adherence self-efficacy among low-literacy patients: development of a pictographic visual analogue scale.** *Health Educ Res* 2005, **20(1)**:24-35.
11. Hibbard JH, Stockard J, Mahoney ER, Tusler M: **Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers.** *Health Serv Res* 2004, **39(4 Pt 1)**:1005-1026.
12. Enwald HP, Huotari ML: **Preventing the obesity epidemic by second generation tailored health communication: an interdisciplinary review.** *J Med Internet Res* **12(2)**:e24.
13. Block G, Sternfeld B, Block CH, Block TJ, Norris J, Hopkins D, Quesenberry CP Jr, Husson G, Clancy HA: **Development of Alive! (A Lifestyle Intervention Via Email), and its effect on health-related quality of life, presenteeism, and other behavioral outcomes: randomized controlled trial.** *J Med Internet Res* 2008, **10(4)**:e43.
14. Olshefsky AM, Zive MM, Scolari R, Zuniga M: **Promoting HIV risk awareness and testing in Latinos living on the U.S.-Mexico border: the Tu No Me Conoces social marketing campaign.** *AIDS Educ Prev* 2007, **19(5)**:422-435.
15. **Healthfinder.gov.** [http://www.healthfinder.gov/].
16. Department of Veterans Affairs: **Caring for Veterans with HIV Disease: Characteristics of Veterans in VA Care, Fiscal Year 2002.** Public Health Strategic Health Care Group, Center for Quality Management in Public Health; 2003.
17. Anaya HD, Hoang T, Golden JF, Goetz MB, Gifford A, Bowman C, Osborn T, Owens DK, Sanders GD, Asch SM: **Improving HIV screening and receipt of results by nurse-initiated streamlined counseling and rapid testing.** *J Gen Intern Med* 2008, **23(6)**:800-807.
18. Goetz MB, Hoang T, Bowman C, Knapp H, Rossman B, Smith R, Anaya H, Osborn T, Gifford AL, Asch SM: **A system-wide intervention to improve HIV testing in the Veterans Health Administration.** *J Gen Intern Med* 2008, **23(8)**:1200-1207.
19. Knapp H, Anaya HD, Feld JE: **Expanding HIV rapid testing via point-of-care paraprofessionals.** *Int J STD AIDS* 2008, **19(9)**:629-632.
20. Owens DK, Sundaram V, Lazzeroni LC, Douglass LR, Tempio P, Holodniy M, Sanders GD, Shadle VM, McWhorter VC, Agoncillo T, *et al*: **HIV testing of at risk patients in a large integrated health care system.** *J Gen Intern Med* 2007, **22(3)**:315-320.
21. Perlin J: **Need for Routine Human Immunodeficiency Virus (HIV) Risk Assessment and Testing (IL 10-2005-017).**Edited by: Affairs V. Washington, DC; 2005:.
22. Dookeran NM, Burgess JF Jr, Bowman CC, Goetz MB, Asch SM, Gifford AL: **HIV screening among substance-abusing veterans in care.** *J Subst Abuse Treat* 2009, **37(3)**:286-291.
23. Nazi KM, Woods SS: **MyHealtheVet PHR: A Description of Users and Patient Portal Use.** *AMIA Annu Symp Proc* 2008, 1182.
24. Fisher JD, Fisher WA: **Changing AIDS-risk behavior.** *Psychol Bull* 1992, **111(3)**:455-474.
25. Kalichman S, Stein JA, Malow R, Averhart C, Devieux J, Jennings T, Prado G, Feaster DJ: **Predicting protected sexual behaviour using the Information-Motivation-Behaviour skills model among adolescent substance abusers in court-ordered treatment.** *Psychol Health Med* 2002, **7(3)**:327-338.
26. Kitzinger J: **Qualitative research. Introducing focus groups.** *BMJ* 1995, **311(7000)**:299-302.
27. Kidd PS, Parshall MB: **Getting the focus and the group: enhancing analytical rigor in focus group research.** *Qual Health Res* 2000, **10(3)**:293-308.
28. Kitzinger J: **The methodology of focus groups: the importance of interaction between research participants.** *Sociology of Health and Illness* 1994, **16(1)**:103-121.
29. Charmaz K: **Constructing Grounded Theory.** Thousand Oaks, CA: Sage; 2006.
30. Zavestoski S, Brown P, McCormick S, Mayer B, D'Ottavi M, Lucove JC: **Patient activism and the struggle for diagnosis: Gulf War illnesses and other medically unexplained physical symptoms in the US.** *Soc Sci Med* 2004, **58(1)**:161-175.
31. Janz NK, Champion VL, Strecher VJ: **The Health Belief Model.** In *Health Behavior and Health Education*.. 3 edition. Edited by: Glanz K, Rimer BK, Lewis FM. San Francisco: John Wiley 2002:.
32. Bauer UE, Johnson TM, Hopkins RS, Brooks RG: **Changes in youth cigarette use and intentions following implementation of a tobacco control program: findings from the Florida Youth Tobacco Survey, 1998-2000.** *JAMA* 2000, **284(6)**:723-728.
33. Ling JC, Franklin BA, Lindsteadt JF, Gearon SA: **Social marketing: its place in public health.** *Annu Rev Public Health* 1992, **13**:341-362.
34. Futterman DC, Peralta L, Rudy BJ, Wolfson S, Guttmacher S, Rogers AS: **The ACCESS (Adolescents Connected to Care, Evaluation, and Special Services) project: social marketing to promote HIV testing to adolescents, methods and first year results from a six city campaign.** *J Adolesc Health* 2001, **29(3 Suppl)**:19-29.
35. Noar SM, Palmgreen P, Chabot M, Dobransky N, Zimmerman RS: **A 10-year systematic review of HIV/AIDS mass communication campaigns: Have we made progress?** *J Health Commun* 2009, **14(1)**:15-42.
36. Lenhart A, Purcell K, Smith A, Zickuhr K: **Social Media and Young Adults.** *Pew Internet & Americal Life Project* Washington, DC: Pew Research Center; 2010.
37. Madden M: **Older Adults and Social Media.** *Pew Internet & Americal Life Project* Washington, DC: Pew Research Center; 2010.
38. Halamka JD, Mandl KD, Tang PC: **Early experiences with personal health records.** *J Am Med Inform Assoc* 2008, **15(1)**:1-7.
39. Sands DZ, Halamka JD: **PatientSite: patient centered communication, services, and access to information.** In *Consumer Informatics: Applications and Strategies in Cyber Health Care.* Edited by: Nelson R, Ball MJ. New York: Springer-Verlag; 2004:.
40. Hess R, Bryce CL, Paone S, Fischer G, McTigue KM, Olshansky E, Zickmund S, Fitzgerald K, Siminerio L: **Exploring challenges and potentials of personal health records in diabetes self-management: implementation and initial assessment.** *Telemed J E Health* 2007, **13(5)**:509-517.
41. Linzer M, Manwell LB, Williams ES, Bobula JA, Brown RL, Varkey AB, Man B, McMurray JE, Maguire A, Horner-Ibler B, *et al*: **Working conditions in**

primary care: physician reactions and care quality. *Ann Intern Med* 2009, **151(1)**:28-36, W26-29.
42. Lakhan SE, Laird C: **Addressing the primary care physician shortage in an evolving medical workforce.** *Int Arch Med* 2009, **2(1)**:14.
43. Craw JA, Gardner LI, Marks G, Rapp RC, Bosshart J, Duffus WA, Rossman A, Coughlin SL, Gruber D, Safford LA, *et al*: **Brief strengths-based case management promotes entry into HIV medical care: results of the antiretroviral treatment access study-II.** *Journal of acquired immune deficiency syndromes (1999)* 2008, **47(5)**:597-606.
44. Gardner LI, Metsch LR, Anderson-Mahoney P, Loughlin AM, del Rio C, Strathdee S, Sansom SL, Siegal HA, Greenberg AE, Holmberg SD: **Efficacy of a brief case management intervention to link recently diagnosed HIV-infected persons to care.** *AIDS (London, England)* 2005, **19(4)**:423-431.
45. Manirankunda L, Loos J, Alou TA, Colebunders R, Nostlinger C: **"It's better not to know": perceived barriers to HIV voluntary counseling and testing among sub-Saharan African migrants in Belgium.** *AIDS Educ Prev* 2009, **21(6)**:582-593.
46. Department of Veterans Affairs Public Health Strategic Health Care Group: **The State of Care for Veterans with HIV/AIDS.** Washington, DC; 2009.
47. Department of Veterans Affairs: **Elimination of Requirement for Prior Signature Consent and Pre and Post-Test Counseling for HIV Testing.** *Federal Register 38 CFR Part 17* Washington DC; 2009.
48. Jain CL, Wyatt CM, Burke R, Sepkowitz K, Begier EM: **Knowledge of the Centers for Disease Control and Prevention's 2006 routine HIV testing recommendations among New York City internal medicine residents.** *AIDS patient care and STDs* 2009, **23(3)**:167-176.
49. **My HealtheVet.** [https://www.myhealth.va.gov/mhv-portal-web/anonymous.portal?_nfpb=true&_nfto=false&_pageLabel=aboutMHVHome].

doi:10.1186/1756-0500-4-295
Cite this article as: McInnes *et al.*: Use of electronic personal health record systems to encourage HIV screening: an exploratory study of patient and provider perspectives. *BMC Research Notes* 2011 **4**:295.

EDITOR'S CHOICE

Qualitative research

A qualitative approach to Bayes' theorem

Mitchell A Medow,[1] Catherine R Lucey[2]

10.1136/ebm-2011-0007

[1]Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts, USA
[2]Internal Medicine, The Ohio State University College of Medicine, Columbus, Ohio, USA

Correspondence to:
Dr Mitchell A Medow
Section of General Internal Medicine, Boston University School of Medicine, Crosstown Center Building, 2nd Floor, 801 Massachusetts Avenue, Boston, MA 02118, USA;
mitchell.medow@bmc.org

Accepted 5 July 2011

Abstract

While decisions made according to Bayes' theorem are the academic normative standard, the theorem is rarely used explicitly in clinical practice. Yet the principles can be followed without intimidating mathematics. To do so, one can first categorise the prior-probability of the disease being tested for as very unlikely (less likely than 10%), unlikely (10–33%), uncertain (34–66%), likely (67–90%) or very likely (more likely than 90%). Usually, for disorders that are very unlikely or very likely, no further testing is needed. If the prior probability is unlikely, uncertain or likely, a test and a Bayesian-inspired update process incorporating the result can help. A positive result of a good test increases the probability of the disorder by one likelihood category (eg, from uncertain to likely) and a negative test decreases the probability by one category. If testing is needed to escape the extremes of likelihood (eg, a very unlikely but particularly dangerous condition or in the circumstance of population screening, or a very likely condition with a particularly noxious treatment), two tests may be needed to achieve. Negative results of tests with sensitivity ≥99% are sufficient to rule-out a diagnosis; positive results of tests with specificity ≥99% are sufficient to rule-in a diagnosis. This method overcomes some common heuristic errors: ignoring the base rate, probability adjustment errors and order effects. The simplicity of the method, while still adhering to the basic principles of Bayes' theorem, has the potential to increase its application in clinical practice.

Bayes' theorem[1] remains the normative standard for diagnosis, but it is often violated in clinical practice. Attempts to simplify its application with diagnostic computer programs,[2 3] nomograms,[4] rulers[5] or internet calculators[6] have not helped to increase its use. Bayes' theorem helps overcome many well-known cognitive errors in diagnosis, such as ignoring the base rate, probability adjustment errors (conservatism, anchoring and adjustment) and order effects.[7] Bayes' theorem and its underlying precepts are introduced early in medical school and medical texts, for example, Chapter 3 of 392 chapters in *Harrison's Principles of Internal Medicine*.[8] Even so, adherence to Bayes' principles is all but absent – low probability diseases are still tested for causing unneeded cost and risk, and high probability diseases are ignored when a single negative test returns.

The basic idea of Bayes' theorem for medical diagnosis is well accepted. A diagnosis is not necessarily confirmed just because a test was positive. Diagnosis is usually not a binary decision (ie, true of false) turning on a single datum, but a dynamic probabilistic assessment. The post-test probability (also called the updated probability, posterior-probability or positive-predictive value) of a diagnosis is dependent on how likely the diagnosis was before the test was done (the pretest probability, also referred to as the prevalence or prior-probability), the test result (positive or negative) and the ability of the test to discriminate between those afflicted and not afflicted with the disease (test characteristics expressed as sensitivity and specificity, or likelihood ratios). A simple formula, Bayes' theorem, combines these elements to produce the post-test probability of the disease. A positive test increases confidence in a diagnosis, but usually does not indicate certainty. Whether this confidence exceeds a treatment (or action) threshold[9] remains a decision for the clinician and patient. Likewise, a negative test decreases confidence in a diagnosis, but rarely rules it out completely. It is up to those involved to decide if further action is warranted.

What if an easy, non-mathematical method to apply these concepts were available? Could the application of Bayes' theorem find its appropriate place in clinical practice and not be relegated to academic exercises for medical students and residents? Could its benefits in clinical practice finally be realised? There is a simple, qualitative or categorical application of Bayes' theorem that might ease the application of Bayes' underlying precepts. The method is based on categorising the pretest probability and handling a small set of probabilistic categories instead of the full spectrum of continuous probabilities, thus eliminating the need for mathematical calculations. In this paper, we first introduce this qualitative method. Then we present the mathematical justification for the method and the conditions under which it holds. Finally we present some special cases that reinforce the method.

Qualitative Bayes' theorem

Bayes' theorem's concepts can be applied using qualitative methods. First one must commit to the pretest probability – how likely the diagnosis is from the start. This probability is expressed categorically – *very unlikely* (less likely than 10%), *unlikely* (between 10% and 33%), *uncertain* (between 34% and 66%), *likely* (between 67% and 90%) or *very likely* (more likely than 90%) (table 1).

If the initial assessment is *very unlikely* or *very likely*, then in most cases it is not worth further testing according to Bayes' theorem – the results would either confirm what is already near certain or it would minimally move the post-test probability in the opposite direction. Either way, the clinician would not normally take additional actions. There are at least two situations where it is still important to proceed with further testing. First, if the diagnosis is *very unlikely* but needs to be ruled out with more certainty, as in a very dangerous disease, the clinician may want to proceed with testing. For example, at what probability is the clinician willing to send a patient home in whom a diagnosis of subarachnoid haemorrhage is being considered? Similarly, if the diagnosis is *very likely*, but

Table 1 Categorical probabilities

Categorical probability	Numerical probability
Very unlikely	Less likely than 10%
Unlikely	Between 10% and 33%
Uncertain	Between 34% and 66%
Likely	Between 67% and 90%
Very likely	More likely than 90%

needs to be ruled in with more certainty, such as when the treatment is especially dangerous or noxious, further testing is indicated. For example, at what probability is the clinician willing to commit a patient with liver disease to a course of anticoagulation with warfarin for a deep venous thrombosis?

If the probability is in one of the intermediate categories, then further testing is appropriate. The clinician may order a test and then interpret the results. A positive result moves the clinician to the next more likely category. A negative result moves the clinician to the next less likely category.

For example, if the clinician is seeing a 35-year-old man in the office who presents with substernal, exertional chest pain that was relieved with rest, the patient has anginal chest pain. His pretest probability of having coronary artery disease (CAD) is *likely*, about 70%.[10] Further testing is warranted and a stress test is ordered. If the result is negative, the diagnosis of CAD is *uncertain* – not absent. If the result is positive, CAD is *very likely*. But, if the patient were a woman, her pretest probability of having CAD is *unlikely* (about 26%). Further testing is also warranted. In this circumstance, if the stress test is negative, the diagnosis of CAD is *very unlikely*. If the test is positive, CAD is *uncertain* – not definitively present.

When does the qualitative approach to applying Bayes' theorem work?

Using the above intuitive cut-offs, and tests with sensitivities and specificities between 80% and 90%, the above procedure is a good approximation to Bayes' theorem.

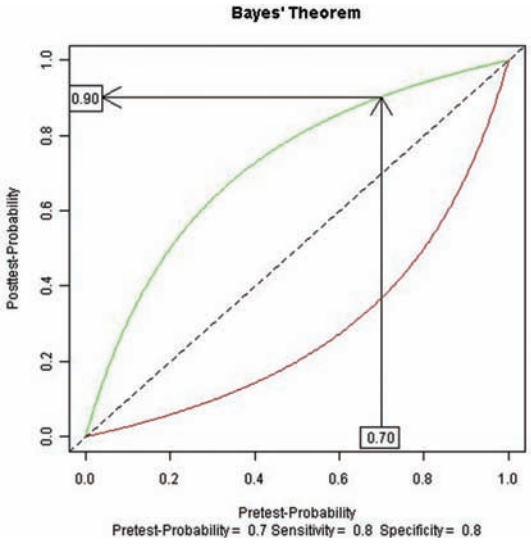


Figure 1 Graphical interpretation of Bayes' theorem.

A graphical approach to Bayes' theorem can demonstrate how the qualitative approximation works (figure 1). Here the horizontal-axis is the pretest probability, the curves represent the relationship between the pretest probability and the post-test probability for a given sensitivity and specificity (80% for each in this example, roughly corresponding to the test characteristics for a nuclear stress test) and the vertical-axis is the post-test probability. The diagonal line is usually included and represents no change in the post-test probability with the test result (ie, the test did not change the clinician's assessment of the probability). Separate curves represent a positive result (green), which increases the post-test probability (ie, is above the diagonal line), and a negative result (red), which reduces the post-test probability. To use Bayes' theorem, one starts on the horizontal-axis at the appropriate pretest probability and draws a vertical line until it intersects the appropriate curve for a positive (green) or negative test (red) result. One then draws a horizontal line to find the appropriate post-test probability on the vertical-axis. Figure 1 exemplifies this for the case above, a man with anginal chest pain and a positive stress test. One first locates 70% on the horizontal-axis, follows the arrow up until it intersects the positive result (green) curve, then follows the arrow horizontally until it intersects the vertical-axis at the post-test probability of 90%. A similar procedure is followed for a negative test, using the red line, giving a post-test probability of 37%.

In the categorical case, all the pretest probabilities between 67% and 90% (the *likely* category) need to be considered while holding the sensitivity and specificity constant at 80%. This only needs to be done for the lower (67%) and the upper limits (90%), see figure 2. For a positive test, the post-test probabilities range between 89% and 97% (arrows). To get the lower limit of post-test probability for a positive test, one follows the arrow from a pretest probability of 67% up until point A in figure 2 (where the arrow intersects the curve representing Bayes' theorem's post-test probability for a positive result), and then reads

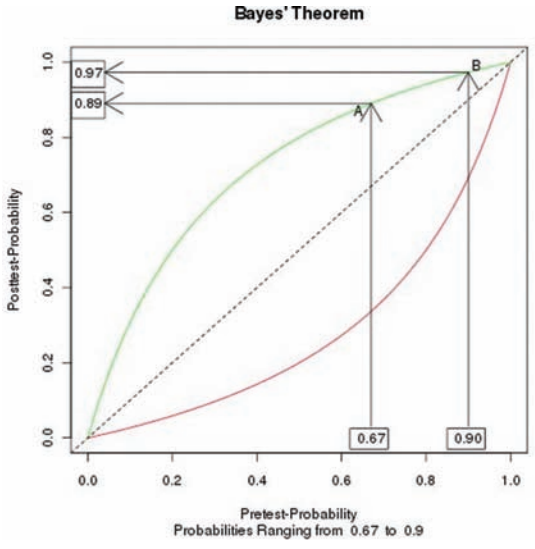


Figure 2 Graphical interpretation of Bayes' theorem for a range of pretest probabilities from 67% to 90% (*likely* category).

the post-test probability (89%) off the vertical-axis. For the upper limit of the post-test probability, one follows the arrow from a pretest probability of 90% up until point B in figure 2, and then finds the post-test probability (97%) on the vertical-axis. This gives the range of post-test probabilities for the *likely* category. For a negative result, the post-test probabilities range between 33% and 69%.

Next, expand the range of sensitivities and specificities from 80% to 90%, representing good tests. Now, instead of a (green) *curve* to represent the relationship between pretest and post-test probabilities, we have a (green) *band* (figure 3). Like before, one follows the arrow from the pretest probability until it firsts meets the band (A in figure 3) to get the lower limit of the post-test probability and until it meets the top of the band (B in figure 3) to get the upper limit of the post-test probability. For the *likely* category examined above, we can see that the post-test probabilities for a positive test now range between 89% and 99% – almost all in the *very likely* category – and for a negative test between 18% and 69% – almost all in the *uncertain or unlikely* categories. The transformation of the pretest probabilities is shown as the purple inverted 'L' in figure 3. The results for all the categories are shown in figure 4 and table 2.

The method is an approximation. It forces a Bayesian inspired analysis on the interpretation of test results and gives results consistent with Bayes' theorem. The approximation is weakest in the two cells with contradictory data (eg, *unlikely* with a *positive* test result and *likely* with a *negative* test result) as expected. The results remain a good approximation even with expanded ranges for test characteristics (sensitivity and specificity). For example, if the range of test characteristics is between 70% and 80%, probability cut-offs of <20%, 20–45%, 46–55%, 56–80% and >80% work well. Since the method is only an approximation and our estimates of pretest probabilities are also poor estimates, just using the cut-offs *likely*, *uncertain* and *unlikely* will suffice.

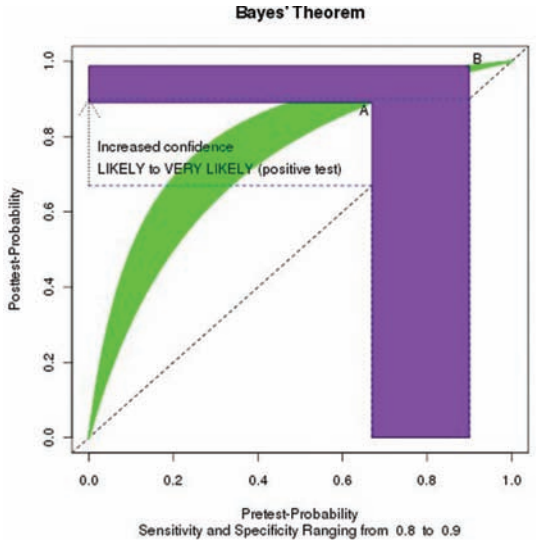


Figure 3 Graphical interpretation of Bayes' theorem for a range of pretest probabilities from 67% to 90% (*likely* category), and sensitivities and specificities ranging between 80% and 90%.

Special cases

Rule-in and rule-out tests – tests in which a single result is capable of near definitively ruling-in or ruling-out a diagnosis – are important. For example, a low brain natriuretic peptide is suitable for ruling-out systolic heart failure. Any test with sensitivity greater than 99% is sufficient to rule-out a diagnosis from even the *likely* category (SnOUT) and any test with specificity greater than 99% is sufficient to rule-in a diagnosis from even the *unlikely* category (SpIN).

In the *very* categories, since the curves are fairly flat in this region (figure 1, for example), two tests might be needed to produce a clinically significant change in the probabilities.

This method does not apply to most screening tests because the pretest probabilities are so low. The method reminds the clinician that a positive result on a screening test is usually not diagnostic, because the change in probabilities is not large enough with a single test. It will usually take two tests to go from *very unlikely* (as target conditions are in the general population) to *very likely* (the final probability a clinician is interested in before undertaking a colectomy, mastectomy or prostatectomy). For example, a 45-year-old woman has a 5-year probability of having breast cancer of about 1%.[11] The sensitivity of routine screening mammography ranges from 71% to 96% and the specificity ranges from 94% to 97%.[12] Using values of 80% for sensitivity and 96% for specificity, a positive test increases the probability to 17%. Using the qualitative categories described herein, the woman's risk of breast cancer would go from *very unlikely* to *unlikely* with the single positive screening test.

Many diseases and tests have appropriate prevalences, and sensitivities and specificities for the tests published, for example, tropinin I for myocardial infarction[13] or urine Chlamydia infection in men.[14]

Summary

In summary, here is a qualitative procedure to follow to approximate the results of a Bayesian diagnostic decision analysis.

1. What is the pretest probability of the disease being considered? Ideally this comes from an evidence-based source. If it is very *likely* (<10–20%) or *very unlikely* (>80–90%), in general, no further testing is needed.
2. One first categorises the pretest probability as *likely, uncertain* or *unlikely*.
3. If the test is positive, the post-test probability increases by one qualitative category (eg, unlikely to uncertain). If the test is negative, the post-test probability decreases by one qualitative category (eg, *unlikely to very unlikely*).
4. This process continues until the clinician is comfortable enough with the confidence in the diagnosis considering the patient's preferences, the risk of the disease and the effects of treatment.
5. Negative tests with sensitivities near 99% can almost certainly rule out a disease, since the post-test sensitivity will be *very unlikely* even if the original pretest probability was *likely*. Similarly, positive tests with

specificities near 99% can almost certainly rule in a disease.

6. If the pretest probability was *very likely* or *very unlikely*, and further testing is indicated, two tests are needed to escape the very categories. This is because the change in the probabilities is small within these categories. Two concordant results are needed to change out of the very categories.

Table 2 Pretest and post-test probabilities for a categorical version of Bayes' theorem

	Test result			
Category	**Positive**		**Negative**	
Unlikely	0.31	0.82	0.01	0.11
Uncertain	0.66	0.95	0.05	0.33
Likely	0.89	0.99	0.18	0.69

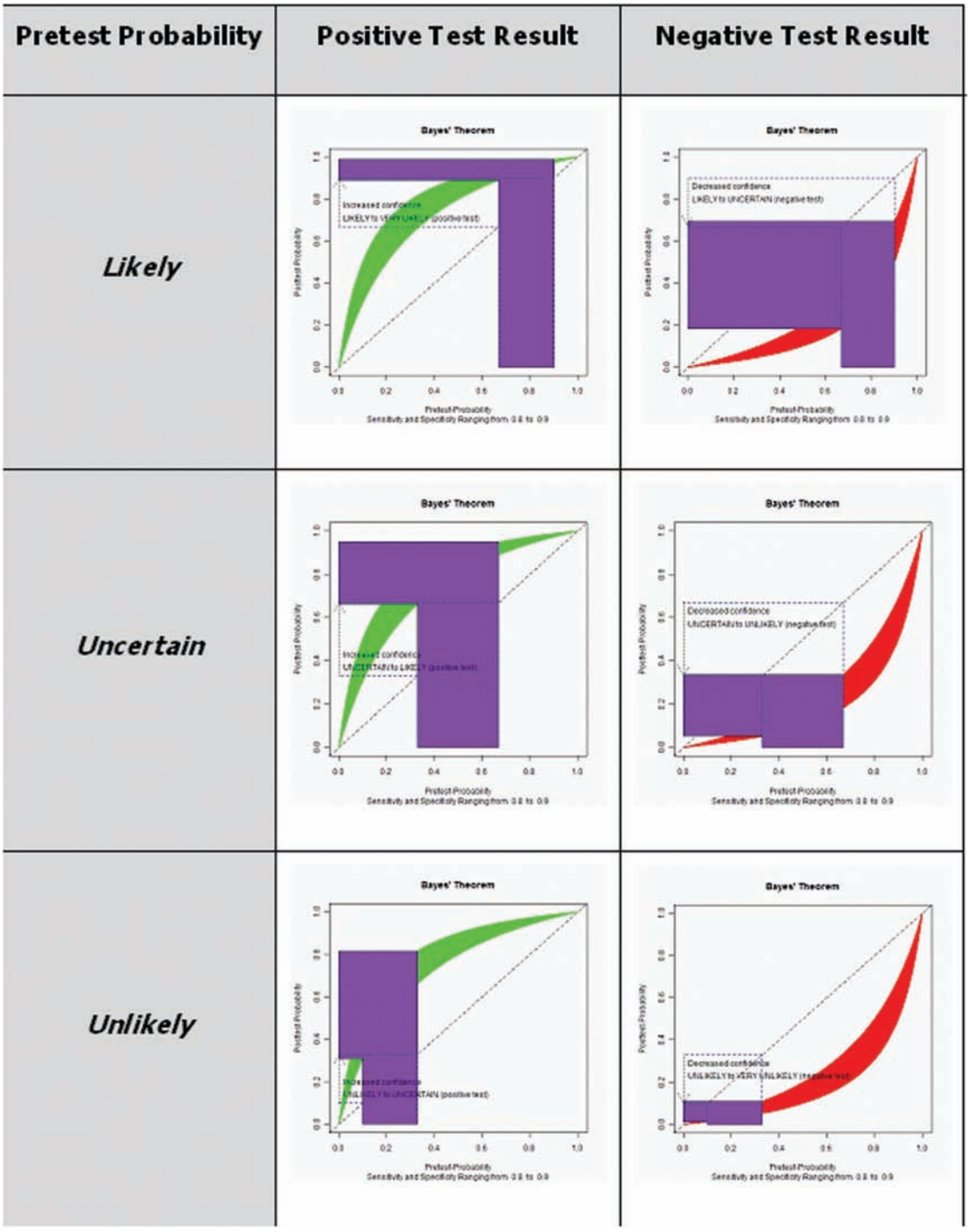


Figure 4 Graphical interpretation of Bayes' theorem for a categorical probabilities, and sensitivities and specificities between 80% and 90%.

Acknowledgments The authors would like to thank Warren Hershman, MD, MPH for his comments.

Competing interests None.

References

1. **Bayes T**, Price M. An Essay Towards Solving a Problem in the Doctrine of Chances. Philosophical Transactions of the Royal Society of London, 1763:370–418.
2. **Horrocks JC**, McCann AP, Staniland JR, *et al*. Computer-aided diagnosis: description of an adaptable system, and operational experience with 2,034 cases. *Br Med J* 1972;**2**:5–9.
3. **Friedman CP**, Elstein AS, Wolf FM, *et al*. Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. *JAMA* 1999;**282**:1851–6.
4. **Fagan TJ**. Letter: nomogram for Bayes theorem. *N Engl J Med* 1975;**293**:257.
5. **Hellmich M**, Lehmacher W. A ruler for interpreting diagnostic test results. *Methods Inf Med* 2005;**44**:124–6.
6. **Schwartz A**. Diagnostic Test Calculator (Version 2010042101). http://araw.mede.uic.edu/cgi-bin/testcalc.pl (accessed 2 February 2011).
7. **Elstein AS**, Schwartz A, Schwarz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ* 2002;**324**:729–32.
8. **Mark DB**. Decision-making in clinical medicine. In: Fauci AS, Braunwald E, Kasper DL, Hauser SL, Longo DL, Jameson JL, Loscalzo J, eds. *Harrison's Principles of Internal Medicine*. 17th edition. Harrison's, NY: McGraw-Hill Professional 2008:2958.
9. **Pauker SG**, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980;**302**:1109–17.
10. **Diamond GA**, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. *N Engl J Med* 1979;**300**:1350–8.
11. **National Cancer Institute**. Breast Cancer Risk Assessment Tool. http://www.cancer.gov/bcrisktool/RiskAssessment.aspx?current_age=45&tage_at_menarche=99&tage_at_first_live_birth=99&ever_had_biopsy=99&previous_biopsies=99&biopsy_with_hyperplasia=99&related_with_breast_cancer=99&race=1 (accessed 23 March 2011).
12. **Cheng T**, Freund KM, Lane S, *et al*. Screening for Breast Cancer. ACP PIER: The Physicians' Information and Education Resource. http://pier.acponline.org/physicians/screening/s369/effectiveness/s369-s3.html (accessed 23 March 2011).
13. **Keller T**, Zeller T, Peetz D, *et al*. Sensitive troponin I assay in early diagnosis of acute myocardial infarction. *N Engl J Med* 2009;**361**:868–77.
14. **Nadala EC**, Goh BT, Magbanua JP, *et al*. Performance evaluation of a new rapid urine test for chlamydia in men: prospective cohort study. *BMJ* 2009;**339**:b2655.

PAIN® 152 (2011) 397–402

PAIN®

www.elsevier.com/locate/pain

Identifying prescription opioid use disorder in primary care: Diagnostic characteristics of the Current Opioid Misuse Measure (COMM)

Ellen C. Meltzer [a], Denis Rybin [b], Richard Saitz [a,c,d], Jeffrey H. Samet [a,e], Sonia L. Schwartz [a], Stephen F. Butler [f], Jane M. Liebschutz [a,e,*]

[a] Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA
[b] Data Coordinating Center, Boston University School of Public Health, Boston, MA, USA
[c] Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA
[d] Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA, USA
[e] Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA
[f] Inflexxion, Inc., Newton, MA, USA

ARTICLE INFO

Article history:
Received 1 December 2009
Received in revised form 14 October 2010
Accepted 3 November 2010

Keywords:
Prescription drug use disorder
Primary care
Diagnosis

ABSTRACT

The Current Opioid Misuse Measure (COMM), a self-report assessment of past-month aberrant medication-related behaviors, has been validated in specialty pain management patients. The performance characteristics of the COMM were evaluated in primary care (PC) patients with chronic pain. It was hypothesized that the COMM could identify patients with prescription drug use disorder (PDD). English-speaking adults awaiting PC visits at an urban, safety-net hospital, who had chronic pain and had received any opioid analgesic prescription in the past year, were administered the COMM. The Composite International Diagnostic Interview served as the "gold standard," using DSM-IV criteria for PDD and other substance use disorders (SUDs). A receiver operating characteristic (ROC) curve demonstrated the COMM's diagnostic test characteristics. Of the 238 participants, 27 (11%) met DSM-IV PDD criteria, whereas 17 (7%) had other SUDs, and 194 (82%) had no disorder. The mean COMM score was higher in those with PDD than among all others (ie, those with other SUDs or no disorder, mean 20.4 [SD 10.8] vs 8.4 [SD 7.5], $P < .0001$). A COMM score of $\geqslant 13$ had a sensitivity of 77% and a specificity of 77% for identifying patients with PDD. The area under the ROC curve was 0.84. For chronic pain patients prescribed opioids, the development of PDD is an undesirable complication. Among PC patients with chronic pain-prescribed prescription opioids, the COMM is a promising tool for identifying those with PDD.

1. Introduction

Multiple studies have shown dramatic increases in prescriptions for opioid analgesics for chronic non-cancer chronic pain in the United States [12,25,31]. Although more Americans use marijuana for pain relief, the number of first time abusers of prescription pain medications has recently exceeded the number of new marijuana users [1]. With the initial diagnosis and management of chronic, non-cancer pain falling largely under the domain of the primary care physician (PCP), many of these doctors report they are not adequately trained to recognize and manage patients at high-risk for, or experiencing, prescription drug use disorder (PDD) [6]. Experts in addiction and pain debate what constitutes PDD in a chronic pain population [15,30]. Although there is some consensus regarding clinical features that patients with PDD typically exhibit, no single "gold standard" exists for diagnosing PDD in primary care (PC) patients with chronic pain [2,4,5,9,29].

Current practice guidelines recommend using the Current Opioid Misuse Measure (COMM) to assess patients who are prescribed opioid therapy [7]. Developed by experts in pain and addiction, the COMM is a patient self-report assessment of past-month aberrant medication-related behaviors, defined as behaviors that are concerning for addiction or taking a medication in a way other than how it is prescribed [5,27]. Aberrant medication-related behaviors may include PDD as well as unintentional misuse, purposeful diversion, or addiction to substances other than pain medication. The COMM validation study was conducted with patients treated in specialty pain management clinics, and a score of 9 or greater was determined to be suggestive of prior 30-day prescription opioid misuse.

The diagnostic capabilities of the COMM have not been evaluated among PC patient populations. Diagnostic tests may perform differently when used in clinical settings other than those in which

* Corresponding author at: Section of General Internal Medicine, Boston University Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA. Tel.: +617 414 3846; fax: +617 414 4676.
E-mail address: jane.liebschutz@bmc.org (J.M. Liebschutz).

0304-3959/$36.00
doi:10.1016/j.pain.2010.11.006

they are validated. The COMM may serve as a practical means of monitoring PC patients treated with opioid therapy for the development of PDD; however, it remains to be determined whether this group of patients can be accurately assessed for PDD with this tool. Using a DSM-IV diagnosis of PDD as a gold standard, the diagnostic performance characteristics of the COMM were evaluated among a sample of PC patients with chronic pain who had received prescription opioids in the past year. The research team chose a broad sample of those at risk for PDD because the clinical challenges are not limited to those using daily current opioids. It was hypothesized that the COMM could identify participants with PDD and could distinguish them from all others. Second, as an exploratory aim, it was hypothesized that the COMM can differentiate those with PDD, some of whom may have a comorbid illicit drug disorder and/or comorbid past year alcohol dependence, from participants with a lone other substance use disorder (SUD) (ie, lone illicit drug disorder and/or past-year alcohol dependence), a prior disorder (PDD or SUD), or no disorder. A receiver operating characteristic (ROC) curve was constructed to evaluate whether the established COMM threshold score of 9 is suggestive of PDD in this patient sample.

2. Methods

2.1. Study design

This was a cross-sectional study of PC patients with chronic pain, defined as lasting for 3 months or longer, completed at the PC clinics of an urban, safety-net, academic medical center [3,18]. The study consisted of 2 parts: an interview with a trained member of the research team, and a subsequent electronic medical record review for abstraction of prescription opioid data to meet entry criteria.

2.2. Setting

Patients waiting for scheduled PC visits were recruited by trained research interviewers. Interviewers were physicians, master's degree-level professionals, college graduates, and college students who underwent 60+ hours of interview training. All participants were approached in the waiting rooms of an academic, urban, safety-net hospital primary care practice [18] Safety-net hospitals in the United States care for poor and vulnerable populations who may be uninsured or underinsured, and includes disproportionate numbers of underrepresented minorities [5].

2.3. Recruitment and enrollment

Between February 2005 and August 2006, a total of 2194 patients were approached, of whom 822 (37.4%) were eligible for the study based on explicit criteria (ie, were 18–60 years of age, spoke English, endorsed pain of at least 3 months' duration, reported use of any analgesic medication (including over-the-counter or prescription) in the prior month, and had a scheduled PC appointment). More than 75% of those eligible (620/822) agreed to participate in the study. Electronic medical record entries from 12 months before study entry were reviewed looking for documentation of an opioid prescription. Standardized chart abstraction forms were used and the electronic medical records were comprehensive. They included notes from all clinic visits, all emergency department records, all inpatient discharge summaries, phone notes, and an institutional prescription database. Patients were eligible for inclusion in this study if they had at least 1 prescription for any of the following opioids in the prior year: butorphanol/stadol; codeine/Tylenol# 2, 3, 4; fentanyl oral/Actiq; fentanyl transdermal (Duragesic); hydrocodone (Vicodin, Norco, Zydone, Maxidone, Lortab, Lorcet, Hydrocet, Co-Gesic, Anexsia); hydromorphone (Dilaudid); meperidine (Demerol); methadone (for pain, not maintenance treatment); morphine-immediate release (MSIR); morphine-extended release (MSContin); nalbuphine (Nubain); oxycodone-immediate release (Percocet, Roxicet, Endocet, Tylox, OxyIR, Roxicodone); Oxycodone-Long acting (Oxycontin); Pentazocine (Talwin); Propoxyphene (Darvon, Darvocet); levorphanol (Levo-Dromoran); and oxymorphone (Numorphan, Opana, Opana ER). Thus, the 238 patients that were prescribed an opioid pain reliever in the prior 12 months were the study sample for this analysis. Informed consent was obtained from eligible patients, and participants were compensated $10. The Boston University Medical Center Institutional Review Board approved the study.

2.4. Measures and key variables

Unless otherwise noted, all variables are obtained from subject interview.

2.5. Study terminology

For the purpose of this study, PDD will be used to describe participants who meet DSM-IV criteria for current (past year) prescription opioid abuse or dependence [2]. During the interview portion, participants were assessed using the Composite International Diagnostic Interview (CIDI) v.2.1 module on Drug Disorders [21]. Using the CIDI, PDD was defined as meeting DSM-IV criteria for current (past-year) prescription opioid abuse or dependence [21]. Criteria for abuse included social, physical, or legal consequences from use. The criteria for dependence included compulsive use, health consequences, and physical dependence (ie, tolerance or withdrawal). Physical dependence alone did not suffice to meet the diagnosis. Participants with PDD could also have comorbid other SUDs.

Other SUD will describe participants who meet DSM-IV criteria for any current (past year) illicit drug abuse or dependence and/or past year alcohol dependence [21]. These were assessed using the CIDI v.2.1 module on Drug Disorders (illicit drugs) and the CIDI-short form (CIDI-SF) for alcohol dependence [21]. Participants with PDD may also have another SUD (ie, comorbid illicit drug disorder and/ or past-year alcohol dependence), but will only be analyzed in the group labeled PDD. Prior drug disorder was defined as meeting DSM-IV criteria for prior (>12 months ago) prescription drug disorder and/or illicit drug disorder [21]. Current alcohol abuse and past alcohol use disorders were not measured using the full CIDI; instead the CIDI-SF was used to reduce respondent time burden. Nicotine dependence was not included in the variable SUD. (Fig. 1). For the main analysis, the participants with Current PDD were compared with all others, which included some with other SUDs. For the exploratory analyses, participants were assigned to one of the following groups: Current PDD, Current Other SUD (have SUD other than PDD), Prior SUD (with or without PDD), and No Lifetime Disorder.

Chronic pain has many different definitions, but experts agree that it is pain that persists for months or years [26]. For the purpose of this study, chronic pain is defined as pain of at least 3 months' duration.

2.6. COMM measure

During the interview portion of the study, each participant was administered a 40-question beta version of the Current Opioid Misuse Measure [5]. Subsequently, the COMM was narrowed down to 17 questions during its validation study [5]. The 17 questions include one newly constructed question. Specifically, question K23 from the beta version, "How often has something happened that

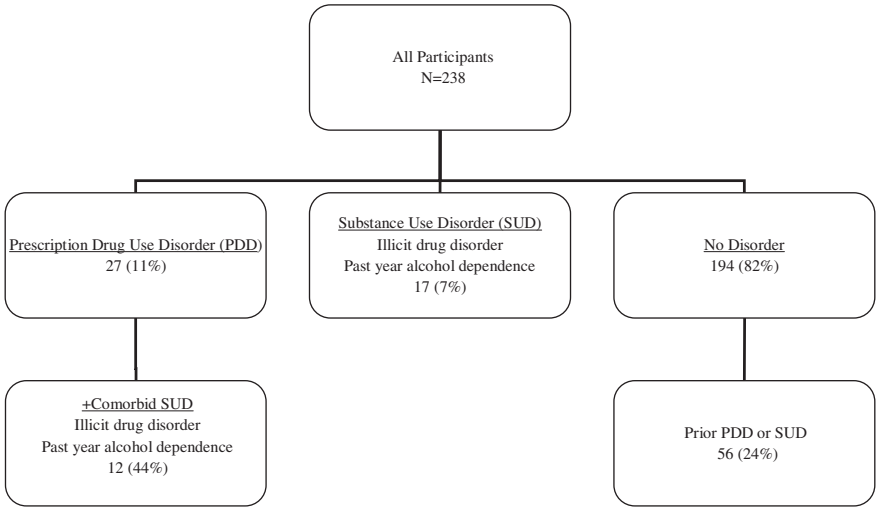


Fig. 1. Study participants and DSM-IV diagnoses.

has worried you about how you're handling your medications?" was separated into two COMM questions: COMM question 10, "How often have you been worried about how you're handling your medications?" and COMM question 11, "How often have others been worried about how you're handling your medications?" For the study presented, participants' scores were calculated based on the 16 questions present in the beta version that were retained in the final COMM questionnaire. Some of the participants' scores would have been higher had not that question been omitted.

2.7. Other variables

The following key variables were examined: (1) sociodemographic factors including age (in years), gender, race/ethnicity (African American/black, Hispanic, white, other), income ($\geqslant$\$20,000 or <\$20,000), employment (unemployed or receiving disability payments vs other), education (less than high school, high school or higher education), marital status (partnered, divorced, single), health insurance (Medicaid/Medicare vs others, including private and uninsured); (2) lifetime post-traumatic stress disorder (PTSD) diagnosis from the CIDI v. 2.1 PTSD module [21]; (3) current Major Depression from the Patient Health Questionnaire (PHQ) for Depression, a 9-item validated measure correlated with past 2-week Major Depression [17]; (4) family history of SUD (single question about first-degree relatives having alcohol or drug problems); (5) current cigarette smoking (taken from the visit closest to the interview date during the electronic medical record review) [18].

2.8. Statistical analysis

Descriptive statistics were calculated using frequencies, means, medians, and standard deviations. To describe the level of opioid medication prescription, we grouped the participants by number of equivalent pills of 5 mg oxycodone (the most common opioid medication prescription), given the plethora of different types of prescriptions, including medication, strength, dosing instructions, and number of fills (original plus any refills). Participants with PDD were compared to all others using *t* tests for continuous data and the Fisher's exact test for categorical data.

As the COMM scores were not distributed normally, all statistical analyses were conducting using both parametric and nonparametric tests of difference. To address the first hypothesis, that the COMM can identify subjects with PDD and distinguish them from all others, the *t* test and Wilcoxon rank-sum test were performed to examine COMM scores by drug disorder groups (PDD versus no disorder) [10]. Both parametric and nonparametric analysis of variance (*F* test and Kruskal–Wallis test) were used to explore the second hypothesis, that the COMM can differentiate participants with PDD from subjects with a single other SUD, a prior disorder (PDD or SUD), or no disorder. As both the parametric and nonparametric tests yielded the same statistically significant results, the mean scores are reported in this paper. Finally, a receiver operating characteristic (ROC) curve was constructed. Using this curve, it was determined whether a threshold score of 9 was suggestive of PDD in this patient sample. The data analysis for this paper was generated using SAS/STAT 9.1 statistical software (SAS Institute, Cary, NC). The Type I error level for all tests was set at 0.05.

3. Results

The demographic characteristics of all 238 participants, stratified by DSM-IV diagnosis of PDD or no PDD are presented (Table 1). Among the entire subject panel, 15% of the subjects received the equivalent of 20 tablets of 5 mg oxycodone in <2 fills, 12.6% received 21–60 tablets in $\leqslant$3 fills, 22.7% received 61–150 tablets

Table 1
Demographic characteristics of participants stratified by PDD diagnosis.

Variable	Current PDD N = 27 (11%) (%)	No PDD = 211 (89%) (%)	*P* value
Mean age (SD)	47.0 (7.9)	46.6 (8.6)	.85
Female gender	15 (56%)	118 (56%)	1.0
Race/ethnicity			
African American	19 (70%)	125 (60%)	.36
Hispanic	(11%)	1 (10%)	
White	(19%)	2 (20%)	
Other	0	21 (10%)	
Education			
<12 y	9 (33%)	62 (29%)	.66
>12 y	8 (67%)	49 (71%)	
Receiving disability payments	16 (59%)	105 (50%)	.42
Lifetime PTSD	13 (48%)	73 (35%)	.20
Current depression (past 2 wk)	17 (63%)	74 (35%)	.006
Current smoking	21 (81%)	97 (47%)	.001
Current alcohol dependence or drug disorder[a]	12 (44%)	17 (8%)	<.001
Mean (SD) COMM score	20.4 (10.8)	8.4 (7.5)	<.001

[a] Other than prescription opioid.

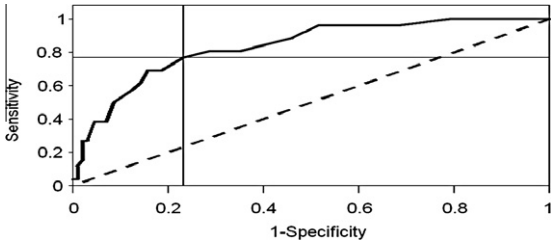


Fig. 2. Receiver operating characteristic curve. Current Opioid Misuse Measure (COMM) prediction score sensitivity and specificity estimates measured against a DSM-IV diagnosis of prescription drug use disorder (PDD). Area under the curve = 0.84 (95% confidence interval, 0.76, 0.91). Diagonal line represents chance prediction.

in ⩽3 fills, and 49.6% received >150 tablets or >3 fills of any amount (eg, 4 prescriptions of 20 tablets each). The majority of those in the last category received >6 fills. Of the participants, 11% (27/238) met DSM-IV criteria for current PDD. There were few differences among the 2 groups of subjects with respect to mean age, distribution of gender and race/ethnicity, and education level attained. The sample had a mean age in the 40s, was largely African American, and the majority had 12 or more years of education. At least 50% of those with PDD and those with no PDD were receiving disability payments, and nearly a third of each group had lifetime post-traumatic stress disorder. Consistent with other studies examining clinical risk factors for PDD [14,18,19], participants with PDD were more likely to experience current depression, to smoke, or to have past-year other drug disorder.

The mean COMM score for those subjects with current PDD, 20.4 (SD 10.8), was significantly higher than those with no current PDD, 8.4 (SD 7.5), $P < .0001$, as was the median COMM score 18.5 vs 7.5, $P < .0001$ (Table 1). Among all participants, COMM scores ranged from 0 to 45. An ROC curve of the COMM data compared to a DSM-IV diagnosis of PDD was constructed (Fig. 2). The area under the curve was 0.84 (95% confidence interval = 0.76, 0.91). Diagnostic performance characteristics across a range of possible COMM scores are presented in Table 2. In this sample, a COMM score of thirteen has the maximum sum of sensitivity and specificity, with a sensitivity of 0.77 and a specificity of 0.77, for identifying participants with DSM-IV PDD. At this value, the positive predictive value (PPV) is 0.30 and the negative predictive value (NPV) is 0.96, and the positive and negative likelihood ratios are 3.31 and 0.30, respectively. This indicates that the probability of having PDD with a positive COMM score is 30%, whereas the probability of not having PDD when the COMM is normal (ie, below the threshold value of 13) is 96% [11].

All 238 subjects were categorized according to whether they met criteria for a DSM-IV diagnosis of current PDD (11%, 27/238), current other SUD (7%, 17/238), or no disorder (82%, 194/238). Participants with no disorder were further categorized based on whether they met criteria for a prior disorder (PDD or other SUD) (24% 56/238) or had no lifetime disorder (58% 138/238) (Fig. 1). Mean COMM scores were calculated for participants with PDD, current other SUD, prior disorder (PDD or SUD), and no lifetime disorder. The mean score for those with PDD remained significantly different from all other groups, including participants with a current other SUD, whereas the other groups did not differ significantly from each other (Table 3).

Table 3
Mean COMM scores for Current PDD vs all others.

Disorder group	Mean (SD) COMM score	Median score	*P* value
Current PDD	20.4 (10.8)	18.5	<.0001
Current other SUD	13.0 (7.4)	12.0	
Prior disorder	9.1 (8.3)	6.0	
No lifetime disorder	7.6 (6.9)	6.0	

4. Discussion

Using a DSM-IV diagnosis of PDD as a gold standard, the diagnostic performance characteristics of the COMM were evaluated among a sample of PC patients with chronic pain who had received prescription opioids in the prior year. The data confirm the hypothesis that the Current Opioid Misuse Measure can distinguish those with a DSM-IV diagnosis of PDD. When compared with the findings presented in the original COMM validation study, conducted among a cohort of patients from specialty pain clinics, the diagnostic characteristics of the COMM seem different in our urban, academic PC patient sample [5]. In that original study, the threshold (cutoff) value used was 9 to detect opioid misuse as defined by a composite measure (not a diagnosis of PDD). In our data, with 2 questions different from the original, a threshold value of 13, rather than 9, maximized the sum of sensitivity and specificity for identifying patients with DSM-IV diagnosis of PDD.

A ROC curve analysis suggests a cutoff point of 13 to maximize the sensitivity and specificity of the COMM within this PC population. The area under the curve of 0.84 implies that the test is good, or moderately accurate, for identifying participants with DSM-IV PDD [13]. As in the original COMM validation study, the selected cutoff score results in greater sensitivity so that few cases of actual PDD are missed [5]. Changing the cutoff to obtain greater specificity limits the number of false-positive results [11]. Individual clinicians, based on the overall prevalence of PDD in their own patient population, may decide to choose a COMM score that maximizes either sensitivity or specificity, rather than the sum of the 2 values [11].

Results also support the exploratory hypothesis that the COMM appears to distinguish patients with PDD, some of whom may have comorbid illicit drug disorders and/or past-year alcohol dependence, from those with a single other SUD, a prior drug disorder (PDD or SUD), or no disorder. The discriminatory capacity of the COMM supports the content validity of the tool [11]. This is

Table 2
COMM Prediction Score vs DSM-IV diagnosis.

COMM score	Sensitivity	Specificity	PPV	NPV	Positive likelihood ratio	Negative likelihood ratio
7	0.961	0.484	0.196	0.989	1.866	0.079
8	0.884	0.540	0.201	0.972	1.924	0.213
9	0.846	0.595	0.215	0.967	2.094	0.258
10	0.807	0.646	0.230	0.962	2.284	0.297
11	0.802	0.681	0.25	0.964	2.538	0.282
12	0.807	0.712	0.269	0.965	2.805	0.270
13	0.769	0.767	0.303	0.962	3.311	0.300
14	0.692	0.813	0.327	0.952	3.704	0.3784
15	0.692	0.843	0.367	0.954	4.421	0.364
16	0.615	0.858	0.363	0.944	4.351	0.447
17	0.576	0.873	0.375	0.940	4.569	0.484

particularly valuable, as patients with other current SUDs or prior disorders (PDD or SUD), while being at higher risk for opioid misuse, may not be currently abusing prescription opioids. Clinicians must appropriately monitor these patients, and the COMM appears to specifically measure PDD [8,20].

The predictive value calculations demonstrate that primary care clinicians can feel fairly confident that patients (in a population with a comparable prevalence of PDD) with a COMM score of less than 13 do not have PDD. However, only 30% of patients with a COMM score of 13 or greater will have PDD. These data reflect the fact that predictive value calculations are affected by the prevalence of disease in a population [11]. Furthermore, as concluded in the original COMM validation study, this tool appears to identify some patients who are not likely to be having problems with their prescription opioids[5]; rather, some of those patients identified as positive will be false positive—ie, patients identified as misusing their medication when they are not. Clinicians are encouraged to practice caution when interpreting the COMM scores and to take into consideration other extenuating circumstances [5,23].

The cutoff COMM score obtained in this study was higher than that obtained in the first published validation study [5]. One possible explanation for this finding can be derived from the fact that a different gold standard was chosen for the current study, using the CIDI to assess for DSM-IV criteria for PDD or other SUDs, whereas the original validation study compared the 40-question beta version of the COMM to the Aberrant Drug Behavior Index [4,5,9]. The Aberrant Drug Behavior Index may measure aberrant medication related behaviors that do not meet criteria for DSM-IV PDD, but are thought by experts to be indicative of prescription opioid "misuse" or "nonmedical use"[15,24,27,30]. In addition, unlike the DSM-IV, the Aberrant Drug Behavior Index will label a patient as having PDD if they use an illicit substance, such as cocaine, while prescribed as an opiate. Using the CIDI as the gold standard permitted a comprehensive assessment of participants for a variety of substance use disorders including PDD, other SUD, and prior disorder (PDD or SUD). Performing this complete analysis focused on use of this instrument for patients at high risk for PDD who may receive opioid analgesic therapy [6,7,19,20,27]. However, primary care clinicians may consider using COMM scores with lower cutoffs as a trigger to discuss potential misuse of the medication in addition to potential PDD.

The COMM was developed as a self-administered questionnaire, and could be incorporated into standard practice for patients chronically prescribed opioids [5]. It requires less than 10 minutes and is easily scored by adding the responses. Ideal timing of the measure (eg, every month, twice yearly) and its utility in combination with treatment contracts, urine toxicology screens, pills counts and prescription monitoring will need to be studied

This study has certain limitations that require consideration. For example, the cross-sectional design does not allow for patients to be followed over time, limiting the types of inferences possible. Specifically, the COMM was only administered once to study participants, so we lack test re-test reliability. In addition, there were a small number of study participants with a DSM-IV diagnosis of current PDD. However, the overall sample size was large enough to produce unambiguous and statistically significant results in each test. These findings do support the need for larger studies in which primary care patients are followed up prospectively, and the COMM is administered repeatedly. Because geography and culture heavily influence use of prescription opioids in clinical and addiction contexts, it is not clear whether these findings are generalizable to areas outside the US or even different primary care populations within the United States.

Another aspect of the study design that should be taken into consideration is the study's reliance, relying on the electronic medical record for data regarding opioid prescriptions. Thus, it might mean that participants who received opioid prescriptions from providers outside of the medical center were excluded from the study. However, by obtaining primary prescription data, recall bias was minimized [22]. Furthermore, implementation of the COMM is oriented toward clinical practices that prescribe opioid therapy, thus providing some assurance that the appropriate patients will ultimately benefit from its use.

The study is also limited by the fact that some participants were not prescribed chronic opioids, and the data analysis did not control for the dose or duration of opioid therapy. Consequently, these results may be less relevant to patients who are prescribed chronic or high-dose opioid therapy. Current guidelines define chronic opioid therapy as "daily or near-daily use of opioids for at least 90 days" [7,16]. It is plausible that some subjects in this study demonstrated behaviors consistent with addiction due to inadequately controlled pain. The extent to which this occurred would have biased the results toward the null hypothesis. Referred to as pseudoaddiction, this preoccupation with opioids often resolves once the pain is adequately controlled [27,28]. For experts in pain and addiction, there is valid concern about patients with pseudoaddiction being inaccurately labeled as PDD [27,28].

Finally, as with any screener, there are always false positives and false negatives. As noted by Butler et al. [5], the COMM is only one source of patient information and should not be used as the sole means of determining whether opioid therapy is appropriate.

5. Conclusions

Among a sample of PC patients with chronic pain had received prescription opioids in the past year, the Current Opioid Misuse Measure (COMM) can identify patients with PDD. Overall, the COMM is a unique clinical tool that demonstrates utility for PC clinicians. Not only does it serve as a validated measure for assessing PDD, but it also provides a means of tracking these behaviors to identify patients at-risk for prescription opioid misuse.

For patients with chronic pain-prescribed opioids, the development of PDD is a serious complication. For primary care physicians treating patients with chronic pain with prescription opioids, the COMM is a promising tool for identifying patients who may have PDD and for helping to confirm that the probability for PDD is low. Future research, in which prospective studies of the COMM are conducted in a variety of PC settings is needed.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Acknowledgments

This study was funded by K23 DA016665 (to J.M.L) and K24 DA022288 from the National Institute of Drug Abuse. Andrew J. Meltzer, MD, provided valuable comments on the manuscript. A pdf of the official COMM scale can be obtained free of charge for clinical use at www.PainEdu.org. Stephen F. Butler, PhD, is Senior Vice President and Chief Science Office at Inflexxion and is one of the creators of the Current Opioid Misuse Measures (COMM). The other authors report no conflicts of interest.

References

[1] Results from the 2006 National Survey on Drug Use and Health: National Findings. Rockville, MD.: Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies; 2006.

[2] American Psychiatric Association, Task Force on DSM-IV. Diagnostic and statistical manual of mental disorders (DSM-IV-TR). Washington, DC: American Psychiatric Association; 2000.

[3] Baxter RJ, Mechanic RE. The status of local health care safety nets. Health Aff (Millwood) 1997;16:7–23.
[4] Butler SF, Budman SH, Fernandez K, Jamison RN. Validation of a screener and opioid assessment measure for patients with chronic pain. Pain 2004;112:65–75.
[5] Butler SF, Budman SH, Fernandez KC, Houle B, Benoit C, Katz N, Jamison RN. Development and validation of the Current Opioid Misuse Measure. Pain 2007;130:144–56.
[6] Califano JA. Under the counter: The diversion and abuse of controlled prescription drugs in the US. The National Center on Addiction and Substance Abuse. Columbia University; 2005.
[7] Chou R, Fanciullo GJ, Fine PG, Adler JA, Ballantyne JC, Davies P, Donovan MI, Fishbain DA, Foley KM, Fudin J, Gilson AM, Kelter A, Mauskop A, O'Connor PG, Passik SD, Pasternak GW, Portenoy RK, Rich BA, Roberts RG, Todd KH, Miaskowski C. Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. J Pain 2009;10:113–30.
[8] Compton P. Should opioid abusers be discharged from opioid-analgesic therapy? Pain Med 2008;9:383–90.
[9] Compton P, Darakjian J, Miotto K. Screening for addiction in patients with chronic pain and "problematic" substance use: evaluation of a pilot assessment tool. J Pain Symptom Manage 1998;16:355–63.
[10] Dawson B, Trapp RG. Basic and clinical biostatistics. New York: Lange Medical Books/McGraw-Hill; 2004.
[11] Fletcher RH, Fletcher SW. Clinical epidemiology: the essentials. Baltimore: Lippincott Williams & Wilkins; 2005.
[12] Gilson AM, Ryan KM, Joranson DE, Dahl JL. A reassessment of trends in the medical use and abuse of opioid analgesics and implications for diversion control: 1997–2002. J Pain Symptom Manage 2004;28:176–88.
[13] Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med 2000;45:23–41.
[14] Ives TJ, Chelminski PR, Hammett-Stabler CA, Malone RM, Perhac JS, Potisek NM, Shilliday BB, DeWalt DA, Pignone MP. Predictors of opioid misuse in patients with chronic pain: a prospective cohort study. BMC Health Serv Res 2006;6:46.
[15] Katz NP, Adams EH, Benneyan JC, Birnbaum HG, Budman SH, Buzzeo RW, Carr DB, Cicero TJ, Gourlay D, Inciardi JA, Joranson DE, Kesslick J, Lande SD. Foundations of opioid risk management. Clin J Pain 2007;23:103–18.
[16] Korff MV, Saunders K, Thomas Ray G, Boudreau D, Campbell C, Merrill J, Sullivan MD, Rutter CM, Silverberg MJ, Banta-Green C, Weisner C. De facto long-term opioid therapy for noncancer pain. Clin J Pain 2008;24:521–7.
[17] Kroenke K. The PHQ-9: a new depression and diagnostic severity measure. Psychiatr Ann 2002;32:509–21.
[18] Liebschutz JM, Saitz R, Weiss RD, Averbuch T, Schwarz S, Meltzer EC, Clagget-Borne E, Cabral HJ, Samet JH. Clinical factors associated with prescription drug use disorder in urban primary care patients with chronic pain. J Pain 2010;11:1047–55.
[19] Michna E, Ross EL, Hynes WL, Nedeljkovic SS, Soumekh S, Janfaza D, Palombi D, Jamison RN. Predicting aberrant drug behavior in patients treated for chronic pain: importance of abuse history. J Pain Symptom Manage 2004;28:250–8.
[20] Prater CD, Zylstra RG, Miller KE. Successful pain management for the recovering addicted patient. Prim Care Companion J Clin Psychiatry 2002;4:125–31.
[21] Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA. The Composite International Diagnostic Interview. An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry 1988;45:1069–77.
[22] Rothman KJ. Epidemiology: an introduction. New York: Oxford University Press; 2002.
[23] Savage SR. Assessment for addiction in pain-treatment settings. Clin J Pain 2002;18:S28–38.
[24] Substance Abuse and Mental Health Services Administration (SAMHSA), Office of Applied Studies. Results from the 2006 National Survey on Drug Use and Health: National Findings (Office of Applied Studies, NSDUH Series H-32, DHHS Publication No. SMA 07-4293); 2007.
[25] Sullivan MD, Edlund MJ, Fan MY, Devries A, Brennan Braden J, Martin BC. Trends in use of opioids for non-cancer pain conditions 2000–2005 in commercial and Medicaid insurance plans: the TROUP study. Pain 2008;138:440–9.
[26] Trescot AM, Boswell MV, Atluri SL, Hansen HC, Deer TR, Abdi S, Jasper JF, Singh V, Jordan AE, Johnson BW, Cicala RS, Dunbar EE, Helm 2nd S, Varley KG, Suchdev PK, Swicegood JR, Calodney AK, Ogoke BA, Minore WS, Manchikanti L. Opioid guidelines in the management of chronic non-cancer pain. Pain Phys 2006;9:1–39.
[27] Weaver M, Schnoll S. Addiction issues in prescribing opioids for chronic nonmalignant pain. J Addict Med 2007;1:10.
[28] Weissman DE, Haddox JD. Opioid pseudoaddiction—an iatrogenic syndrome. Pain 1989;36:363–6.
[29] Wu SM, Compton P, Bolus R, Schieffer B, Pham Q, Baria A, Van Vort W, Davis F, Shekelle P, Naliboff BD. The addiction behaviors checklist: validation of a new clinician-based measure of inappropriate opioid use in chronic pain. J Pain Symptom Manage 2006;32:342–51.
[30] Zacny J, Bigelow G, Compton P, Foley K, Iguchi M, Sannerud C. College on Problems of Drug Dependence taskforce on prescription opioid non-medical use and abuse: position statement. Drug Alcohol Depend 2003;69:215–32.
[31] Zerzan JT, Morden NE, Soumerai S, Ross-Degnan D, Roughead E, Zang F, Simoni-Wastila L, Sullivan SD. Trends and geographic variation of opiate medication use in state Medicaid fee-for-service programs. Med Care 2006;44:1005–10.

Journal of Aging and Health
23(1) 86–111

DOI: 10.1177/0898264310383156
http://jah.sagepub.com

Older Adults' Inpatient and Emergency Department Utilization for Ambulatory-Care-Sensitive Conditions: Relationship With Alcohol Consumption

Elizabeth S. Levy Merrick, PhD, MSW[1], Dominic Hodgkin, PhD[1], Deborah W. Garnick, ScD[1], Constance M. Horgan, ScD[1], Lee Panas, MS[1], Marian Ryan, RRT, MPH, CHES[1], Frederic C. Blow, PhD[2], and Richard Saitz, MD, MPH[3,4,5]

Abstract

Objective: This study examined the relationship between drinking that exceeds guideline-recommended limits and acute-care utilization for ambulatory-care-sensitive conditions (ACSCs) by older Medicare beneficiaries. **Method:** This secondary data analysis used the 2001-2006 Medicare Current Beneficiary Survey (unweighted n = 5,570 community dwelling, past-year drinkers, 65 years and older). Self-reported alcohol consumption

[1]Brandeis University, Waltham, MA
[2]University of Michigan, Ann Arbor
[3]Boston University School of Medicine, MA
[4]Boston Medical Center, MA
[5]Boston University School of Public Health, MA

Corresponding Author:
Elizabeth S. Levy Merrick, PhD, MSW, Senior Scientist, Institute for Behavioral Health, Heller School for Social Policy and Management, Brandeis University, Mail Stop 035, 415 South St., Waltham, MA 02454
Email: merrick@brandeis.edu

(categorized as within guidelines, exceeding monthly but not daily limits, or heavy episodic) and covariates were used to predict ACSC hospitalization, emergency department visit not resulting in admission, and emergency department visit that did result in admission. **Results:** Heavy episodic drinking was significantly associated with higher likelihood of an ACSC emergency department visit not resulting in admission (adjusted odds ratio = 1.91, 95% CI: 1.11-3.30; $p < .05$). Drinking pattern was not significant for other ACSC measures. **Discussion:** Results partially support the hypothesis that excessive drinking may be related to ACSC acute-care utilization among older adults, suggesting increased risk of lower quality outpatient care.

Keywords

older adults, alcohol, ambulatory-care-sensitive conditions, health care utilization, quality of care

Introduction

Older adults frequently have unhealthy drinking patterns, ranging from risky drinking (which incurs increased risk of adverse consequences) to alcohol disorders (Saitz, 2003). Recent national prevalence estimates for adults aged 65 and older range from 9% drinking in excess of national guidelines (Merrick, Horgan, et al., 2008) to gender-specific rates of 13% and 8% with at-risk use and 14% and 3% with binge drinking (five or more drinks on same occasion in past 30 days) for men and women, respectively (Blazer & Wu, 2009). Among the problems associated with excessive drinking is the possibility of increased risk for inadequate medical care. This study investigated the relationship between excessive drinking and acute-care utilization for ambulatory-care-sensitive conditions (ACSCs).

Excessive drinking could be connected to inadequate medical care in several ways. Patients with excessive drinking may underuse routine care including primary care (Ford, Trestman, Tennen, & Allen, 2005; Girard, Partridge, Becker, & Bock, 2004; Kunz, 1997; Rice & Duncan, 1995; Rice et al., 2000), for reasons including concerns about stigma or financial barriers to care. Excessive alcohol consumption could reflect generalized self-neglect of health (Blow, Brockmann, & Barry, 2004; Hazelton, Sterns, & Chisholm, 2003). In any case, the result may be lower use of recommended services or delays in seeking care until problems are more severe. At the same time, providers sometimes have attitudinal barriers toward patients with alcohol problems (Anderson et al., 2004; Deehan, Templeton, Taylor, Drummond, &

Strang, 1998; Freidman, McCullough, Chin, & Saitz, 2000; Kaner et al., 2009), which could potentially result in differential treatment. These patients may be more difficult to treat, or scarce clinical time may be taken up with competing demands.

Although the mechanisms are not well understood, there is some evidence of lower quality medical care for persons with mental disorders, and in some cases substance use disorders specifically, in disparate clinical contexts (Clark, Weir, Ouellette, Zhang, & Baxter, 2009; Desai, Rosenheck, Druss, & Perlin, 2002; Druss, Rask, & Katon, 2008; Rathore, Wang, Druss, Masoudi, & Krumholz, 2008). Older adults who drink excessively according to various criteria are less likely to receive preventive medical care (Merrick, Hodgkin, et al., 2008; Moore et al., 2001; Ozminkowski et al., 2006).

However, acute-care utilization for ACSC has seldom been explored in relation to excessive drinking. ACSCs are conditions for which "timely and effective outpatient care can help to reduce the risks of hospitalization by either preventing the onset of an illness or condition, controlling an acute episodic illness or condition, or managing a chronic disease or condition" (Billings et al., 1993, p. 163). Congestive heart failure (CHF), diabetes, asthma, and chronic obstructive pulmonary disease (COPD) are examples of ACSCs. Measuring hospitalization for ACSCs is one approach to examining access to adequate outpatient care. Numerous ACSC indicators have been developed, including the widely used Agency for Healthcare Research and Quality (AHRQ) Prevention Quality Indicators (PQIs). These are designed to identify admissions that evidence suggests could have been avoided, at least in part, through better access to high-quality outpatient care (AHRQ, 2008). Researchers have considered which ACSC indicators are most appropriate for an older adult population (McCall, Brody, Mobley, & Subramanian, 2004). The concept and diagnosis specifications have more recently been extended to emergency department visits (Logan, Riley, & Barker, 2008; McCall et al., 2004).

The relationship between alcohol use and ACSC inpatient and emergency department utilization is important to investigate in older adults, who experience adverse health events at high rates and face special issues with alcohol. The National Institute on Alcohol Abuse and Alcoholism (NIAAA) defines risky drinking amounts for those 65 years and older as more than seven drinks per week or more than three drinks on any single day (NIAAA, 2005). Similarly, the American Geriatrics Society defines risky drinking as, on average, no more than 1 drink per day, 7 drinks per week, or 3 drinks on heavier drinking occasions (American Geriatrics Society, 2006). Exceeding these limits is associated with interpersonal and functioning problems for

elders (Moos, Brennan, Schutte, & Moos, 2004), who have higher sensitivity and impaired ability to metabolize alcohol (Saitz, 2003). Limits may be lower or abstinence may be best for older adults who use medications that interact with alcohol or have medical problems that can be adversely affected by alcohol consumption.

Many studies have examined the occurrence and predictors of ACSC hospitalization (and, more recently, ACSC emergency department visits), but few have examined the role of alcohol consumption. Li and colleagues examined New York state hospital discharge data for adults aged 20 to 64 years (Li, Glance, Cai, & Mukamel, 2008). Inpatients with mental disorders (including substance use disorders) were significantly more likely than other admitted inpatients to have been admitted for an ACSC and also had longer average length of stay. Another study with a mixed-age inpatient sample found that binge drinking (five or more alcoholic beverages per occasion during the past 30 days) was predictive of preventable hospitalization (Arozullah et al., 2006).

Other factors found to be associated with ACSC hospitalization include health status, comorbidity, and functional limitations (Culler, Parchman, & Przybylski, 1998; Niefeld et al., 2003); income (Billings, Anderson, & Newman, 1996; Billings et al., 1993; Blustein, Hanson, & Shea, 1998); health insurance status (Weissman, Gatsonis, & Epstein, 1992; Zeng et al., 2006); mental disorders (Bynum et al., 2004; Himelhoch, Weller, Wu, Anderson, & Cooper, 2004); access to regular primary care or community health centers (Culler et al., 1998; Epstein, 2001; Falik, Needleman, Wells, & Korb, 2001); continuity of care (Gill & Mainous, 1998); and race and ethnicity (Culler et al., 1998; Friedman & Basu, 2004; Laditka, Laditka, & Mastanduno, 2003). McCall et al., using the 1999 Medicare Current Beneficiary Survey, found that prior ACSC hospitalization, comorbidity, and health status were among the significant predictors for some ACSCs among older beneficiaries (McCall et al., 2004). No variables for alcohol use or mental disorders were included. For ACSC emergency department use, barriers to primary care may contribute to higher utilization for ACSCs among Black adults and Medicaid patients (Oster & Bindman, 2003). Among older adults, emergency department and observation stay utilization increased between 1992 and 2000 for 10 of 11 ACSCs studied, including CHF, pneumonia, and cellulitis (McCall et al., 2004).

The current study aimed to help address the lack of research in this area by focusing on the relationship between self-reported alcohol consumption and ACSC acute-care utilization in a nationally representative sample of older adults. We hypothesized that drinking that exceeded guideline-recommended

limits, particularly heavy episodic drinking, would be positively related to ACSC hospitalization and emergency department use.

Method

Data and Sample

The primary data source was the Medicare Current Beneficiary Survey (MCBS) for 2001-2006. The MCBS is an ongoing survey of a representative national sample of the Medicare population by the Centers for Medicare and Medicaid Services. The sample is selected using a stratified, multistage probability design to represent the national Medicare population (Centers for Medicare and Medicaid Services, n.d.). Sample weights are provided to achieve nationally representative estimates, in this analysis for the continuously enrolled Medicare population. Beneficiaries sampled from Medicare enrollment files (or proxies) are interviewed three times a year including in-person, computer-assisted interviewing. There is a 4-year rotating panel design. Beneficiaries were randomly selected according to age strata from a nationally representative set of 107 geographic primary sampling units, with oversampling of the disabled (age <65) and those aged 65 and older. Normalized sampling weights were assigned to represent the population. The survey content includes sociodemographics, health, and functional status, and utilization. The 2001, 2003, and 2005 MCBS included items regarding alcohol consumption as well as data for covariates (detailed in the Measures section; Centers for Medicare and Medicaid Services, n.d.). Subjects' Medicare claims were linked to survey data for this analysis. We used data from 2001, 2003, and 2005 for baseline characteristics including drinking and covariates, and from 2002, 2004, and 2006 to identify hospitalization and emergency department services for ACSCs. This approach seeks to avoid reverse causality bias between ACSC use and other variables. To maximize sample size, we included beneficiaries present in the MCBS and continuously enrolled for any two consecutive years: 2001 and 2002, 2003 and 2004, or 2005 and 2006.

The analytic sample identification started with 20,482 community-dwelling beneficiaries who were 65 years or older. Of these, 13,612 were nondrinkers and another 129 were missing alcohol data and were excluded. This analysis focused on persons who reported drinking alcohol in a typical month in the past year because nondrinkers would constitute an especially heterogeneous group including lifetime abstainers and those who quit due to health problems, and these differences would be unobservable. Health maintenance organization enrollees were excluded because their claims were not available ($n = 1,171$).

Our final study sample included 5,570 persons representing a weighted *N* of 15,128,450. Bivariate comparison totals varied due to item-missing data (all <4%). The logistic regression sample consisted of 5,046 individuals.

Measures

Dependent Variables: Measures of Acute-Care Utilization for ACSCs.

The three dependent variables were binary measures of any ACSC hospitalization, ACSC emergency department visit not resulting in inpatient admission (thus not overlapping with ACSC hospitalization), and ACSC emergency department visit that did result in admission. We identified ACSCs by applying the specifications from two sources. First, following AHRQ specifications, we used all 13 relevant PQIs (excluding low birth weight): diabetes short-term complications, perforated appendix, diabetes long-term complications, COPD, hypertension, CHF, dehydration, bacterial pneumonia, urinary tract infection, angina without procedure, uncontrolled diabetes, adult asthma, and rate of lower extremity amputation among patients with diabetes (AHRQ, 2008). Second, we adopted 15 indicators identified by McCall et al. (2004) through literature review and deemed appropriate by clinical experts for application to older adults: asthma/COPD, cellulitis, CHF, dehydration, diabetes, hypertension, hypoglycemia, hypokalemia, influenza, urinary tract infection, malnutrition, perforated or bleeding ulcer, pneumonia, seizure disorder, and severe ear/nose/throat infection. The specifications were adopted from the original source and an expanded version published in a later ACSC study (Zeng et al., 2006).

There is substantial overlap between the two indicator lists. If a utilization event qualified in terms of either source, we counted it as an ACSC event. For sensitivity analysis purposes, we examined frequencies for each list of indicators separately and found that the differences were quite small in magnitude. We identified hospitalizations with an ACSC principal diagnosis. Although the PQI set was developed for application to inpatient hospitalization, we follow previous research that has extended the ACSC concept and diagnostic specifications to emergency department visits (Logan et al., 2008; McCall et al., 2004).

We identified emergency department visits based on procedure and location codes. For emergency department utilization that did not result in hospital admission, we used the visit primary diagnosis. Emergency department visits that resulted in admission were necessarily identified through the inpatient file, and the primary admitting diagnosis was attributed to the preceding emergency department visit.

Alcohol Consumption Variables. The 2001, 2003, and 2005 MCBS included three alcohol consumption items. Quantity and frequency were ascertained by asking, "Please think about a typical month in the past year. On how many days did you drink any type of alcoholic beverage? On those days that you drank alcohol, how many drinks did you have?" Heavy episodic drinking was assessed by asking, "Please think about a typical month in the past year. On how many days did you have four or more drinks in a single day?" Alcoholic beverages were described as including "liquor such as whiskey or gin, mixed drinks, wine, beer, and any other type of alcoholic beverage."

To assess unhealthy drinking in terms of consuming risky amounts of alcohol (regardless of whether alcohol consequences or disorders were present), we defined alcohol measures reflecting two parameters in the NIAAA and American Geriatrics Society guidelines (American Geriatrics Society Clinical Practice Committee, n.d.). First, to be consistent with the weekly guideline we defined *exceeding monthly limits* as more than 30 drinks per typical month. (A total of 42 respondents reporting 31 drinks per month whose responses were clearly based on a 31-day month were coded as negative as the items did not specify standardized number of days per month.) Second, we constructed a heavy episodic drinking variable indicating whether an individual reported four or more drinks in any single day during a typical month in the past year, according to either drinking quantity item.

We categorized respondents into three mutually exclusive categories: within-guidelines drinkers (not exceeding the monthly limit or the three-drink, single-day limit), drinkers who exceeded the monthly limit but not the single-day limit, and heavy episodic drinkers who exceeded the single-day drinking limit, with or without exceeding the monthly limit. For secondary analyses, we calculated a continuous measure of drinks per month, based on the quantity–frequency responses. To address nonlinearity, we included a squared term in the regressions.

Covariates. Covariates were selected that previous research found to affect health care utilization.

Sociodemographic variables. We included gender, race, Hispanic ethnicity, annual household income, age, education, marital status, region, and residence in a metropolitan area. Living arrangement was not included due to high correlation with marital status.

Health status variables. We controlled for health status by utilizing DxCG (diagnostic cost group) risk adjustment software that uses sex, age, and diagnosis codes from claims to construct a continuous measure of relative risk of health care resource use (DxCG, 2009). Compared to other illness burden indices or scales, the DxCG score contains higher specificity related to the

individual's clinical profile in projecting future health care costs and estimating an individual's care management needs (Zhao, Ash, Ellis, & Slaughter, 2002; Zhao et al., 2001). Thus, it may be used as a proxy for health status in that higher DxCG risk scores denote higher health care resource use risk and presumably poorer health (Wang et al., 2000). A value of 1 indicates the individual's predicted cost equals the population average for all persons with Medicare claims; higher values indicate higher than average predicted costs. For bivariate analyses, we created categories: no claims or claims not indicative of significant health risk (DxCG score < 0.1), claims indicative of lower than average health risk (0.1 ≤ DxCG score ≤ 1), and claims indicative of higher than average health risk (DxCG score > 1). For logistic regression models, we used the continuous measure; increasing scores indicate higher risk of health care resource use (poorer health status). We also included dichotomous variables for current smoking and for presence of a chronic disease explicitly related to ACSC indicators. We used diagnosis codes from the Centers for Medicare and Medicaid Services' Chronic Condition Data Warehouse (COPD, diabetes, CHF; Buccaneer Computer Systems and Services, 2009) and the AHRQ PQI specifications (hypertension, asthma); two outpatient claims or one inpatient claim during the baseline year were required for all but CHF in which one claim of any type was required.

We controlled for functional status using a modified Katz Index of Independence in Activities of Daily Living (ADL) variable constructed from survey data (Katz, Ford, Moskowitz, Jackson, & Jaffe, 1963; Shelkey & Wallace, 1999). Respondents were asked whether they had trouble or needed assistance with six ADLs: bathing, dressing, transferring, toileting, continence, or feeding. If no difficulty was indicated, that activity received a score of 1. The resulting variable reflects a 7-position scale (0-6) of the number of independent ADLs.

Two mental health variables were used. First, a self-reported depression variable was created. Respondents were asked, "In the past 12 months, how much of the time did you feel sad, blue, or depressed?" (all, most, some, little, or none of the time), and "In the past 12 months, have you had 2 weeks or more when you lost interest or pleasure in things that you usually cared about or enjoyed?" (yes, no). Respondents who answered "all" or "most of the time" to the first question and/or "yes" to the second question were categorized as having self-reported depression. This approximates the modified PHQ-2 validated for older adults (Li, Friedman, Conwell, & Fiscella, 2007). Second, a dichotomous variable indicating behavioral health diagnosis was constructed based on the presence of one inpatient claim or two outpatient claims with an International Classification of Diseases (ninth edition; ICD-9) mental health

or substance use disorder diagnosis, excluding alcohol disorders. We excluded alcohol diagnosis from our main models as we wanted to observe the full effect of self-reported drinking, but we included this in alternate versions. We found that although in bivariate analyses the presence of a claim with an alcohol disorder diagnosis was associated with significantly higher rates of ACSC acute-care utilization, it was not statistically significant in a multivariate context.

Access variables. We included measures on self-reported trouble getting needed care during the past 12 months, private supplemental insurance, and Medicaid coverage. We also included a measure for living in a county designated as a full-county primary care health professional shortage area according to Health Resources and Services Administration data (Health Resources and Services Administration, n.d.). In addition, we constructed variables indicating type of usual care: no particular "medical person or clinic" usually seen when respondent is sick or for advice, except for hospital emergency room or walk-in urgent care center; primary care physician (family practice, general practice, geriatrics, internal medicine); or nonprimary care physician. We constructed a variable indicating a prior-year ACSC event. Due to multicollinearity, it was not included in final models. When we tested its inclusion, this variable was positive and significant but did not affect the direction, magnitude, or significance of alcohol variables.

Statistical Analysis

Results are weighted estimates that represent the continuously enrolled, community-dwelling, non-HMO, elderly Medicare population of current drinkers. Chi-square tests were used to assess bivariate differences involving drinking categories; chi-square statistics were corrected for the survey design and converted to *F* statistics. The continuous drinks per month variable was significantly skewed, and bivariate testing accounted for skewness. We conducted logistic regression analyses to model occurrence of an ACSC hospitalization or emergency room visit as a function of alcohol consumption (drinking category) and covariates. We also constructed a continuous measure of alcohol consumption—number of drinks per month—and included in a separate set of logistic regression models. We explored the use of an instrumental variables approach to address any potential bias due to unobserved differences between drinking groups. Among several exogenous variables (e.g., Sunday ban on alcohol sales, beer tax) that have been previously used as instruments for alcohol use, none were found for this sample of older drinkers that met the assumptions required for

instrumental variables analysis (Wooldridge, 2006) and were robust to specific geographic inclusion (notably Puerto Rico).

Due to the complex sampling design, using procedures that assumed equal probability of selection would likely lead to underestimating standard errors (Cohen, 1997; Lemeshow et al., 1998). The SVY: LOGIT procedure of the statistical package STATA version 9.0 was used to more accurately determine the statistical significance of observed differences (STATA Corporation, n.d.).

Results

As shown in Table 1, the weighted sample was predominantly White (93.0%), male (55.2%), and married (65.4%). Only 16.2% were more than 80 years old. More than half had some education beyond high school. Three quarters (75.2%) reported within-guidelines alcohol consumption in a typical month in the past year. Ten percent exceeded the monthly limit only, and 14.8% reported heavy episodic drinking. Drinking patterns varied significantly by sociodemographic, health, and access variables.

Number of drinking days, drinks per drinking day, and drinks per month (weighted) varied significantly by drinking category ($p < .01$; data not shown). As noted earlier, the continuous measure of drinks per month was constructed from quantity-frequency variables, which were not harmonized with the separate heavy episodic drinking survey item. Among within-guidelines drinkers, the mean number of drinking days during a typical month was 9.1 (standard error *[SE]* 0.2) with median of 3.7, and the mean number of drinks per drinking day was 1.2 (*SE* 0.01) with median of 0.8. The mean number of drinks per month was 10.3 (*SE* 0.2) and the median was 4.8. For drinkers exceeding monthly limits only, the mean number of drinking days during a typical month was 27.7 (*SE* 0.2) with a median of 29.3, and the mean number of drinks per drinking day was 2.2 (*SE* 0.02) with a median of 2.0. The mean number of drinks per month was 59.9 (*SE* 0.6) and the median was 58.9. For heavy episodic drinkers, the mean number of drinking days was 18.7 (*SE* 0.5) with a median of 19.8, and the mean number of drinks per drinking day was 3.8 (*SE* 0.1) with a median of 3.0. The mean number of drinks per month was 67.9 (*SE* 3.7) and the median was 50.0. The mean number of days with heavy drinking was 7.3 (*SE* 0.4) with a median of 2.9.

Overall, 3.5% had an ACSC hospitalization during the year, 2.1% had an ACSC emergency department visit that did not result in inpatient admission, and 2.1% had an emergency department visit that resulted in admission (Table 2). Drinking category was not significantly related to ACSC hospitalization in bivariate analyses. Drinking category was significantly associated

Table 1. Sample Description and Distribution Among Drinking Categories

	Weighted row percentage			
	Total % of sample (weighted column %)	Drinks within guidelines	Exceeds monthly limit only	Heavy episodic drinking
Unweighted *n*	5,570	4,215	543	812
Weighted *N*	15,128,450	11,373,766	1,516,640	2,238,044
Total	100.0	75.2	10.0	14.8
Sociodemographics				
Gender**				
Female	44.8	85.1	7.6	7.3
Male	55.2	67.1	12.0	20.9
Age**				
65 to 70 years	39.1	70.5	10.9	18.6
71 to 80 years	44.7	76.6	9.5	13.9
≥81 years	16.2	82.6	9.2	8.1
Hispanic**	4.3	70.9	5.6	23.5
Non-Hispanic	95.7	75.4	10.2	14.4
Race**				
White	93.0	75.5	10.4	14.2
African American and Other[a]	7.0	72.5	4.8	22.7
Education**				
<High school diploma	15.7	70.3	7.3	22.4
High school graduate	27.6	75.6	9.3	15.1
Some college/voc/tech	27.9	77.0	10.2	12.8
College degree	28.8	75.6	12.2	12.2
Annual household income**				
<US$25,000	37.5	75.7	7.4	16.9
US$25,000 to US$40,000	27.1	76.2	8.9	15.0
>US$40,000	35.4	73.5	13.1	13.4
Marital status**				
Married	65.4	74.7	10.5	14.8
Widowed	22.8	79.9	8.4	11.7
Divorced, separated, single	11.8	69.0	10.2	20.8

(continued)

Table 1. (continued)

	Weighted row percentage			
	Total % of sample (weighted column %)	Drinks within guidelines	Exceeds monthly limit only	Heavy episodic drinking
Metropolitan area**				
Lives in metro area	81.0	76.2	10.0	13.8
Nonmetropolitan area	19.0	71.0	10.0	19.0
Region**				
Northeast	21.1	76.7	11.4	12.0
Midwest	27.4	77.5	7.3	15.2
West	19.3	72.9	13.4	13.7
South and other[a]	32.2	73.7	9.4	16.9
Health and functional status				
Relative health risk**				
No risk indication	36.1	72.8	9.5	17.7
<average risk	45.5	76.6	10.2	13.1
>average risk	18.4	76.3	10.5	13.2
Self-reported depression				
Depressed most of time	8.4	72.3	9.4	18.2
Not depressed most of time	91.6	75.4	10.1	14.5
Mental health diagnosis				
Has mental health diagnosis	7.3	77.9	9.1	13.0
No mental health diagnosis	92.7	75.0	10.1	14.9
Current smoker**				
Is current smoker	13.0	57.6	13.0	29.4
Not current smoker	87.0	77.8	9.6	12.6
Functional status				
Independence in 6 ADLs	70.1	74.4	10.3	15.2
Independence in 5 ADLs	22.7	76.4	10.0	13.6
Independence in 0 to 4 ADLs	7.3	79.3	6.5	14.1
Chronic disease (selected)*				
Has chronic disease	56.3	77.0	9.4	13.6

(continued)

Table 1. (continued)

	Weighted row percentage			
	Total % of sample (weighted column %)	Drinks within guidelines	Exceeds monthly limit only	Heavy episodic drinking
No chronic disease	43.7	73.8	10.5	15.7
Access variables				
Lives in PCSA**				
Lives in PCSA	10.1	68.6	11.1	20.4
Not in PCSA	89.9	76.0	9.9	14.2
Trouble getting care				
Had trouble	2.7	80.0	6.3	13.7
Did not have trouble	97.3	75.0	10.1	14.8
Usual source of care**				
No usual source (except emergency/urgent care)	5.4	63.3	10.6	26.1
Primary care	85.6	76.2	10.1	13.7
Nonprimary care	9.0	76.2	9.6	14.1
Private insurance**				
Has private insurance	82.9	76.3	10.4	13.3
No private insurance	17.1	69.7	8.2	22.1
Medicaid coverage**				
Has Medicaid coverage	4.5	68.4	5.0	26.6
No Medicaid coverage	95.5	75.5	10.3	14.2

Note: PCSA = primary care shortage area. *Ns* for bivariate comparisons varied due to item-missing data, <3.9% for all variables. Percentages are for nonmissing data.
a. Some values not displayed separately due to containing unweighted cell $n < 11$.
*$p < .05$. **$p < .01$.

with ACSC emergency department visits not resulting in admission (1.9% for within-guidelines drinkers, 2.8% for those exceeding monthly limits only, and 3.1% for heavy episodic drinkers, $p < .05$). Drinking category was not significantly related to ACSC emergency department visits that did result in inpatient admission. The continuous measure of drinks per month was not significantly associated with ACSC measures in bivariate tests.

Table 2. Utilization for ACSC, By Drinking Pattern

Unweighted n = 5,570, weighted N = 15, 128, 450	Weighted percentage with any ACSC Hospitalization	Weighted percentage with any ACSC emergency department visit—no inpatient admission	Weighted percentage with any ACSC emergency department visit—resulted in inpatient admission
Overall percentage	3.5	2.1	2.1
Drinking category		*	
Within guidelines	3.5	1.9	2.2
Drinks over monthly limit only	3.7	2.8	2.3
Heavy episodic	3.8	3.1	1.8
Sociodemographics			
Gender			
Female	3.2	2.1	1.9
Male	3.8	2.2	2.3
Age	**	**	**
65 to 70 years	2.3	1.2	1.6
71 to 80 years	3.8	2.5	2.0
≥81 years	5.8	3.5	3.8
Ethnicity			
Hispanic	3.5	2.3	3.0
Non-Hispanic	5.0	2.1	2.1
Race	**		
White	3.3	2.1	2.0
African American or Other[a]	6.8	2.6	3.4
Education	**		**
<High school diploma	5.5	2.7	3.7
High school graduate	3.4	2.4	1.5
Some college/voc/tech	3.5	1.8	2.0
College degree	2.6	1.9	2.0
Annual household income	**		**
<US$25,000	5.0	2.2	3.3
US$25,000 to US$40,000	3.3	2.2	1.6
>US$40,000	2.3	2.1	1.2

(continued)

Table 2. (continued)

Unweighted $n = 5,570$, weighted $N = 15,128,450$	Weighted percentage with any ACSC Hospitalization	ACSC emergency department visit—no inpatient admission	ACSC emergency department visit—resulted in inpatient admission
Marital status	**		
Married	2.9	2.0	1.8
Widowed	5.2	2.7	2.8
Divorced, separated, single	3.7	1.8	2.4
Metropolitan area			
In metropolitan area	3.6	2.0	2.3
Not in metro area	3.3	2.7	1.4
Region	*		**
Northeast	2.9	2.4	2.1
Midwest	3.5	2.2	1.5
West	2.5	2.2	1.5
South and other[a]	4.6	1.9	3.0
Health and functional status			
Relative health risk	**	**	**
No risk indication	1.2	1.0	0.8
<average risk	3.1	1.9	1.7
>average risk	9.2	4.8	5.8
Self-reported depression		*	*
Depressed most of time	4.7	3.7	3.5
Not depressed most of time	3.4	2.0	2.0
Mental health diagnosis	**		*
Mental health diagnosis	6.4	3.0	3.7
No mental health diagnosis	3.3	2.1	2.0
Current smoker	**		
Current smoker	4.9	2.0	2.9
Not current smoker	3.3	2.2	2.0
Functional status	**	**	**
Independence in 6 ADLs	2.9	1.9	1.7

(continued)

Table 2. (continued)

Unweighted $n = 5,570$, weighted $N = 15, 128, 450$	Weighted percentage with any ACSC Hospitalization	ACSC emergency department visit—no inpatient admission	ACSC emergency department visit—resulted in inpatient admission
Independence in 5 ADLs	4.4	2.1	2.4
Independence in 0 to 4 ADLs	7.9	4.8	5.9
Chronic disease (selected)	**	**	**
Has chronic disease	5.6	3.1	3.2
Does not have chronic disease	1.9	1.4	1.3
Access variables			
Primary care shortage area			
Lives in PCSA	4.2	2.8	2.3
Not in PCSA	3.4	2.1	2.1
Trouble getting care	4.9	1.7	3.4
No trouble getting care	3.5	2.2	2.1
Usual source of care			
No usual source except emergency/urgent care	4.1	1.3	1.5
Primary care	3.5	2.2	2.1
Nonprimary care	3.1	2.3	3.0
Private insurance		*	
Private insurance	3.4	2.3	2.0
No private insurance	4.2	1.3	2.6
Medicaid coverage	**		*
Had Medicaid coverage	7.2	1.9	4.4
No Medicaid coverage	3.4	2.1	2.0

Note: ACSC = Ambulatory-care-sensitive condition; ADLs = activities of daily living.
Weighted sample used; significance based on chi-square tests, corrected for survey design.
N for each bivariate comparison varied slightly due to item-missing data, <4% for all variables.
Percentages shown are for nonmissing data.
a. Values combined to avoid unweighted cell sizes < 11.
*$p < .05$. **$p < .01$. ACSC measure varies significantly by independent variable.

The most common ACSCs for hospitalization were CHF (0.9%), bacterial pneumonia (0.8%), and asthma/COPD (0.7%; data not shown). For emergency department visits that did not result in admission, urinary tract infections (0.4%), asthma/COPD (0.3%), and cellulitis (0.3%) were most common. For emergency department visits that resulted in admission, asthma/COPD (0.4%), CHF (0.4%), and bacterial pneumonia (0.4%) were most common. Although the numbers within each specific ACSC were small, thus limiting statistical power to identify significant differences by drinking pattern, there was a significant bivariate relationship between specific ACSC and drinking pattern in several cases. For example, there was a significant relationship between hypertension and drinking pattern across all types of utilization ($p < .01$), with heavy episodic drinkers having the highest utilization. Urinary tract infection and cellulitis were also significantly associated with drinking pattern ($p < .01$), with persons who exceeded monthly guidelines only showing the highest rates of emergency department visits that did not result in admission.

In the logistic regression model predicting ACSC hospitalization, drinking variables were not significant. Being over 80 years of age, African American, having higher relative health risk score, independence in 0 to 4 ADLs, and chronic disease were associated with higher likelihood of an ACSC hospitalization; living in the West region of the country was associated with lower likelihood relative to the South (Table 3). In the model predicting any ACSC emergency department visit that did not result in inpatient admission, heavy episodic drinking was significantly associated with higher likelihood (adjusted odds ratio = 1.91, 95% confidence interval: 1.11-3.30, $p < .05$). Other factors associated with greater likelihood of utilization included older age, greater relative health risk, independence in 0 to 4 ADLs, and chronic disease. Finally, in the model predicting any ACSC emergency department visit that resulted in inpatient admission, drinking variables were not significant. Greater relative health risk and independence in 0 to 4 ADLs were associated with higher likelihood of this type of utilization. Being a high school graduate (relative to having less than a high school education), annual household income of more than US$40,000 per year, and living in the Midwest or West (relative to South) were associated with lower likelihoods.

In another set of alternate logistic regression analyses using a continuous measure of drinks per month as well as a squared term to address nonlinearity, there was no significant effect on ACSC hospitalization or emergency department utilization resulting in admission (data not shown). However, for emergency department utilization that did not result in admission, there was a significant, positive relationship (odds ratio = 1.01, $p < .05$)

Table 3. Logistic Regression Results (Weighted): Predictors of Hospitalization and Emergency Department Visits for ACSCs

Unweighted n = 5,052; Weighted N = 13,677,112	Odds ratio (95% confidence interval)		
	Any ACSC hospitalization	Any ACSC emergency department visit—no inpatient admission	Any ACSC emergency department visit—resulted in inpatient admission
Drinking category (ref: drinks within guidelines)			
Drinks over monthly limit only	1.14 (0.63-2.07)	1.53 (0.80-2.94)	0.91 (0.46-1.83)
Heavy episodic	0.89 (0.53-1.48)	1.91 (1.11-3.30)*	0.70 (0.36-1.32)
Sociodemographics			
Female	0.74 (0.52-1.07)	0.95 (0.56-1.60)	0.70 (0.47-1.05)
Age (ref: 65-70 years)			
71 to 80 years	1.33 (0.86-2.04)	1.84 (1.05-3.25)**	1.04 (0.62-1.74)
≥81 years	1.77 (1.11-2.81)*	2.41 (1.30-4.47)**	1.54 (0.91-2.61)
Hispanic	1.27 (0.46-3.54)	0.97 (0.30-3.10)	0.98 (0.25-3.80)
Race (ref: White)			
African American	2.38 (1.23-4.60)*	1.98 (0.86-4.56)	1.38 (0.65-2.96)
Other	0.90 (0.33-2.47)	1.59 (0.43-5.90)	0.66 (0.20-2.22)
Education (ref: <high school diploma)			
High school graduate	0.81 (0.51-1.29)	0.96 (0.58-1.61)	0.54 (0.30-1.00)*
Some college/ voc/tech	0.83 (0.53-1.32)	0.79 (0.43-1.44)	0.79 (0.47-1.32)
College degree	0.77 (0.44-1.35)	0.78 (0.41-1.46)	0.95 (0.51-1.75)
Annual household income (ref: <US$25,000)			
US$25,000 to US$40,000	0.93 (0.59-1.48)	1.24 (0.75-2.05)	0.58 (0.32-1.04)
>US$40,000	0.74 (0.46-1.18)	1.37 (0.80-2.34)	0.41 (0.23-0.73)**
Marital status (ref: married)			
Widowed	1.39 (0.98-1.97)	1.11 (0.64-1.93)	1.01 (0.63-1.61)
Divorced, separated, single	1.05 (0.64-1.71)	1.26 (0.59-2.68)	0.94 (0.48-1.2)
Metropolitan area	0.94 (0.63-1.40)	0.75 (0.45-1.24)	1.37 (0.75-2.49)

(continued)

Table 3. (continued)

Unweighted n = 5,052; Weighted N = 13,677,112	Odds ratio (95% confidence interval)		
	Any ACSC hospitalization	Any ACSC emergency department visit—no inpatient admission	Any ACSC emergency department visit—resulted in inpatient admission
Region (ref: South)			
Northeast	0.68 (0.43-1.05)	1.28 (0.76-2.16)	0.68 (0.41-1.12)
Midwest	0.79 (0.55-1.14)	1.17 (0.69-1.98)	0.44 (0.28-0.70)**
West	0.55 (0.33-0.92)*	0.99 (0.60-1.64)	0.48 (0.25-0.90)*
Other	0.83 (0.18-3.86)	1.56 (0.29-8.44)	1.06 (0.13-8.34)
Health and functional status			
Relative health risk	1.46 (1.31-1.62)**	1.22 (1.08-1.38)**	1.46 (1.29-1.64)**
Depressed most of time	0.97 (0.55-1.70)	1.44 (0.83-2.52)	1.23 (0.62-2.42)
Mental health diagnosis	1.0 (0.60-1.72)	0.97 (0.48-1.94)	0.87 (0.43-1.78)
Current smoker	1.40 (0.83-2.35)	0.81 (0.40-1.63)	1.54 (0.81-2.93)
Functional status (ref: Independence in 6 ADLs)			
Independence in 5 ADLs	1.21 (0.85-1.72)	1.08 (0.64-1.80)	1.28 (0.77-2.11)
Independence in 0 to 4 ADLs	1.1 (1.09-3.00)*	2.20 (1.19-4.04)*	2.20 (1.26-3.81)**
Chronic disease	1.89 (1.33-2.69)**	1.64 (1.03-2.62)*	1.48 (0.97-2.27)
Access variables			
Lives in primary care shortage area	0.86 (0.36-2.04)	1.15 (0.43-3.05)	0.57 (0.13-2.45)
Trouble getting care	1.07 (0.40-2.86)	0.27 (0.07-1.05)	0.88 (0.30-2.54)
Usual source of care (ref: primary care)			
None (except emergency/ urgent care)	1.31 (0.62-2.77)	1.01 (0.38-2.74)	0.51 (0.17-1.52)
Nonprimary care	0.68 (0.39-1.18)	1.29 (0.77-2.17)	0.88 (0.44-1.76)
Private insurance	1.22 (0.68-2.19)	1.91 (0.88-4.14)	1.06 (0.55-2.01)
Medicaid coverage	1.24 (0.61-2.51)	0.84 (0.29-2.41)	0.93 (0.38-2.26)
Baseline year 2001	1.10 (0.80-1.52)	1.25 (0.80-1.97)	1.00 (0.64-1.57)

Note: ACSC = Ambulatory-care-sensitive condition; ADLs = activities of daily living.
*$p < .05$. **$p < .01$.

Discussion

Heavy episodic drinking was a predictor of ACSC emergency department use that did not result in admission, but drinking variables did not predict ACSC hospitalization or emergency department visits resulting in hospital admission. Thus, we found partial support for the hypothesis that older adults whose drinking exceeded guideline-recommended limits would be at greater risk of ACSC acute-care utilization. This is consistent with prior studies that found evidence for lower quality of medical or preventive care for persons with substance abuse problems defined in various ways (Arozullah et al., 2006; Clark et al., 2009; Desai et al., 2002; Li et al., 2008; Merrick, Horgan, et al., 2008; Ozminkowski et al., 2006) and extends this line of inquiry by finding this connection between older adults' self-reported drinking and ACSC emergency department use. As ACSC severity level for emergency department visits not resulting in admission is likely lower, our findings suggest that people with heavy episodic drinking may experience some deficiencies in access to or quality of outpatient care but in ways that do not affect the most severe outcomes. Heavy episodic drinkers may be using the emergency department for issues that should normally have been addressed earlier in primary care but are not so far advanced as to require hospitalization. Small numbers greatly inhibited our ability to examine specific ACSCs. However, we did find that for some ACSCs (notably hypertension) that can be worsened by excessive drinking utilization varied by drinking pattern with heavy episodic drinkers more likely to have ACSC utilization.

The fact that this relationship occurs only for those with heavy episodic drinking rather than those who exceed monthly limits only echoes findings from a previous study on receipt of preventive services (Merrick, Hodgkin, et al., 2008). The level of drinking represented in the group exceeding monthly limits only is not extremely high, with most respondents indicating consumption of two drinks per day most days of the month. Older adults whose drinking exceeds monthly limits only may reflect a population without current impairment affecting use of outpatient services that prevents ACSC acute-care utilization. Heavy episodic drinkers may be more likely to have alcohol disorders, compared to persons exceeding monthly guidelines only, some of whom who may be continuing to drink at levels that were acceptable for younger ages.

The alternate, exploratory analysis we conducted using drinks per month as the key explanatory variable, when we accounted for nonlinearity, yielded results similar to the results for heavy episodic drinking. Future research to

investigate the potentially complex influences of drinking pattern as well as drinking quantity in more detail would be fruitful. Future work should also investigate the causal mechanisms underlying the association we identified between episodic heavy drinking (or greater number of drinks per month) and ACSC emergency department visits not resulting in admission. Identifying the provider, patient, and system roles in barriers to care for this group will be useful.

Study limitations include possible underreporting of alcohol consumption, although self-reported alcohol consumption is in general considered to be as accurate as other drinking measures (Babor, Steinberg, Anton, & Del Boca, 2000). Relatively small sample size limited statistical power to detect small differences. As noted earlier, it is possible that unobserved differences between drinking groups could create bias, but we were unable to identify satisfactory instruments for our sample and measures. However, one main possible source of bias in this type of analysis is health status differences, and the multivariate models we used included multiple health and functional status variables, somewhat reducing risk this bias. Furthermore, we focused on past-year drinkers, avoiding the bias that could have resulted from including a large, heterogeneous group of nondrinkers, many of whom might have quit drinking due to ill health. Previous research on effects of drinking on acute-care utilization (but not ACSC specifically) that used instrumental variables found no evidence of bias in a different nationally representative sample of older adults (Balsa, Homer, Fleming, & French, 2008). Another limitation is the imprecision of the continuous drinking measures for heavy episodic drinkers in our secondary analyses. This is due to lack of harmonization across alcohol variables and lack of information on specific number of drinks consumed on heavy drinking days. We also note that the application of ACSC indicators to emergency department utilization is a more recent extension of the concept and specifications originally applied to hospitalization.

The study findings add to the understanding of the full range of risks associated with drinking that exceeds recommended guidelines for older adults. They may be useful to outpatient and emergency room providers in raising awareness of this issue, which in turn may heighten vigilance and effective interventions. This may contribute to efforts to both address alcohol issues and reduce acute-care utilization for ACSCs in older adults. The findings also provide a useful contribution in analyzing the effects of a relevant factor largely absent from the large body of research on predictors of ACSC hospitalization and emergency department utilization.

Acknowledgment

The authors thank Grant Ritter for statistical consultation and Michele Hutcheon for manuscript preparation.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

The authors disclosed that they received the following support for their research and/or authorship of this article: This study was funded by the National Institute on Alcohol Abuse and Alcoholism Grant 5R21AA015746.

References

Agency for Healthcare Research and Quality. (2008). *Prevention quality indicators: Technical specifications*. Rockville, MD: Author. Available from http://www.qualityindicators.ahrq.gov/pqi_download.htm

American Geriatrics Society. (2006). Position paper: Alcohol use disorders in older adults. *Annals of Long-Term Care, 14*, 23-26.

Anderson, P., Kaner, E., Wutzke, S., Funk, M., Heather, N., Wensing, M., et al. (2004). Attitudes and managing alcohol problems in general practice: An interaction analysis based on findings from a WHO collaborative study. *Alcohol Alcohol, 39*, 351-356.

Arozullah, A. M., Lee, S. Y., Khan, T., Kurup, S., Ryan, J., Bonner, M., et al. (2006). The roles of low literacy and social support in predicting the preventability of hospital admission. *Journal of General Internal Medicine, 21*, 140-145.

Babor, T. F., Steinberg, K., Anton, R., & Del Boca, F. (2000). Talk is cheap: Measuring drinking outcomes in clinical trials. *Journal of Studies on Alcohol and Drugs, 61*, 55-63.

Balsa, A. I., Homer, J. F., Fleming, M. F., & French, M. T. (2008). Alcohol consumption and health among elders. *The Gerontologist, 48*, 622-636.

Billings, J., Anderson, G. M., & Newman, L. S. (1996). Recent findings on preventable hospitalizations. *Health Affairs (Millwood), 15*, 239-249.

Billings, J., Zeitel, L., Lukomnik, J., Carey, T. S., Blank, A. E., & Newman, L. (1993). Impact of socioeconomic status on hospital use in New York City. *Health Affairs (Millwood), 12*, 162-173.

Blazer, D. G., & Wu, L. T. (2009). The epidemiology of at-risk and binge drinking among middle-aged and elderly community adults: National Survey on Drug Use and Health. *American Journal of Psychiatry, 166*, 1162-1169.

Blow, F. C., Brockmann, L. M., & Barry, K. L. (2004). Role of alcohol in late-life suicide. *Alcoholism: Clinical and Experimental Research, 28*(5 Suppl.), 48S-56S.

Blustein, J., Hanson, K., & Shea, S. (1998). Preventable hospitalizations and socioeconomic status. *Health Affairs (Millwood), 17*, 177-189.

Buccaneer Computer Systems and Services. (2009). *Chronic condition data warehouse user manual* (Version 1.5). Baltimore: Center for Medicare and Medicaid Services.

Bynum, J. P., Rabins, P. V., Weller, W., Niefeld, M., Anderson, G. F., & Wu, A. W. (2004). The relationship between a dementia diagnosis, chronic illness, Medicare expenditures, and hospital use. *Journal of the American Geriatrics Society, 52*, 187-194.

Centers for Medicare and Medicaid Services. (n.d.). *Medicare current beneficiary survey*. Available from http://www.cms.gov/mcbs

Clark, R. E., Weir, S., Ouellette, R. A., Zhang, J., & Baxter, J. D. (2009). Beyond health plans: Behavioral health disorders and quality of diabetes and asthma care for Medicaid beneficiaries. *Medical Care, 47*, 545-552.

Cohen, S. B. (1997). An evaluation of alternative PC-based software packages developed for the analysis of complex survey data. *Journal of the American Statistical Association, 51*, 285-292.

Culler, S. D., Parchman, M. L., & Przybylski, M. (1998). Factors related to potentially preventable hospitalizations among the elderly. *Medical Care, 36*, 804-817.

Deehan, A., Templeton, L., Taylor, C., Drummond, C., & Strang, J. (1998). Low detection rates, negative attitudes and the failure to meet the "Health of the Nation" alcohol targets: Findings from a national survey of GPs in England and Wales. *Drug and Alcohol Review, 17*, 249-258.

Desai, M. M., Rosenheck, R. A., Druss, B. G., & Perlin, J. B. (2002). Mental disorders and quality of diabetes care in the veterans health administration. *American Journal of Psychiatry, 159*, 1584-1590.

Druss, B. G., Rask, K., & Katon, W. J. (2008). Major depression, depression treatment and quality of primary medical care. *General Hospital Psychiatry, 30*, 20-25.

DxCG. (2009). *Products and services, RiskSmart Stand Alone*. Retrieved October 27, 2009, from http://www.dxcg.com/products-services/risksmart-stand-alone.asp

Epstein, A. J. (2001). The role of public clinics in preventable hospitalizations among vulnerable populations. *Health Services Research, 36*, 405-420.

Falik, M., Needleman, J., Wells, B. L., & Korb, J. (2001). Ambulatory care sensitive hospitalizations and emergency visits: Experiences of Medicaid patients using federally qualified health centers. *Medical Care, 39*, 551-561.

Ford, J. D., Trestman, R. L., Tennen, H., & Allen, S. (2005). Relationship of anxiety, depression and alcohol use disorders to persistent high utilization and potentially problematic under-utilization of primary medical care. *Social Science & Medicine, 61*, 1618-1625.

Freidman, P. D., McCullough, D., Chin, M. H., & Saitz, R. (2000). Screening and intervention for alcohol problems: A national survey of primary care physicians and psychiatrists. *Journal of General Internal Medicine, 15*, 84-91.

Friedman, B., & Basu, J. (2004). The rate and cost of hospital readmissions for preventable conditions. *Medical Care Research and Review, 61*, 225-240.

Gill, J. M., & Mainous, A. G., III. (1998). The role of provider continuity in preventing hospitalizations. *Archives of Family Medicine, 7*, 352-357.

Girard, D. D., Partridge, R. A., Becker, B., & Bock, B. (2004). Alcohol and tobacco use in the elder emergency department patient: Assessment of rates and medical care utilization. *Academic Emergency Medicine, 11*, 378-382.

Hazelton, L. D., Sterns, G. L., & Chisholm, T. (2003). Decision-making capacity and alcohol abuse: Clinical and ethical considerations in personal care choices. *General Hospital Psychiatry, 25*, 130-135.

Health Resources and Services Administration. (n.d.). *Primary care health professional shortage areas*. Available from http://datawarehouse.hrsa.gov/hpsadetail.aspx

Himelhoch, S., Weller, W. E., Wu, A. W., Anderson, G. F., & Cooper, L. A. (2004). Chronic medical illness, depression, and use of acute medical services among Medicare beneficiaries. *Medical Care, 42*, 512-521.

Kaner, E., Bland, M., Cassidy, P., Coulton, S., Deluca, P., Drummond, C., et al. (2009). Screening and brief interventions for hazardous and harmful alcohol use in primary care: A cluster randomised controlled trial protocol. *BMC Public Health, 9*, 287.

Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., & Jaffe, M. W. (1963). Studies of illness in the aged. The index of ADL: A standardized measure of biological and psychosocial function. *JAMA, 185*, 914-919.

Kunz, J. L. (1997). Alcohol use and reported visits to health professionals: An exploratory study. *Journal of Studies on Alcohol and Drugs, 58*, 474-479.

Laditka, J. N., Laditka, S. B., & Mastanduno, M. P. (2003). Hospital utilization for ambulatory care sensitive conditions: Health outcome disparities associated with race and ethnicity. *Social Science & Medicine, 57*, 1429-1441.

Lemeshow, S., Letenneur, L., Dartigues, J. F., Lafont, S., Orgogozo, J. M., & Commenges, D. (1998). Illustration of analysis taking into account complex survey considerations: The association between wine consumption and dementia in the PAQUID study. Personnes Ages Quid. *American Journal of Epidemiology, 148*, 298-306.

Li, C., Friedman, B., Conwell, Y., & Fiscella, K. (2007). Validity of the Patient Health Questionnaire 2 (PHQ-2) in identifying major depression in older people. *Journal of the American Geriatrics Society, 55*, 596-602.

Li, Y., Glance, L. G., Cai, X., & Mukamel, D. B. (2008). Mental illness and hospitalization for ambulatory care sensitive medical conditions. *Medical Care, 46*, 1249-1256.

Logan, J. E., Riley, A. W., & Barker, L. E. (2008). Parental mental and pain-related health and pediatric ambulatory care sensitive emergency department visits and hospitalizations. *Health Services Research, 43*, 656-674.

McCall, N., Brody, E., Mobley, L., & Subramanian, S. (2004, June). *Investigation of increasing rates of hospitalization for ambulatory care sensitive conditions among Medicare Fee-for-Service beneficiaries* (Final report RTI Project Number 08686.001.002, CMS Contract No. 500-00-0029). Research Triangle Park, NC: RTI International.

Merrick, E. L., Hodgkin, D., Garnick, D. W., Horgan, C. M., Panas, L., Ryan, M., et al. (2008). Unhealthy drinking patterns and receipt of preventive medical services by older adults. *Journal of General Internal Medicine, 23*, 1741-1748.

Merrick, E. L., Horgan, C. M., Hodgkin, D., Garnick, D. W., Houghton, S. F., Panas, L., et al. (2008). Unhealthy drinking patterns in older adults: Prevalence and associated characteristics. *Journal of the American Geriatrics Society, 56*, 214-223.

Moore, A. A., Morgenstern, H., Harawa, N. T., Fielding, J. E., Higa, J., & Beck, J. C. (2001). Are older hazardous and harmful drinkers less likely to participate in health-related behaviors and practices as compared with nonhazardous drinkers? *Journal of the American Geriatrics Society, 49*, 421-430.

Moos, R. H., Brennan, P. L., Schutte, K. K., & Moos, B. S. (2004). High-risk alcohol consumption and late-life alcohol use problems. *American Journal of Public Health, 94*, 1985-1991.

National Institute on Alcohol Abuse and Alcoholism. (2005). *Helping patients who drink too much: A clinician's guide*. Retrieved June 5, 2008, from http://pubs.niaaa.nih.gov/publications/Practitioner/CliniciansGuide2005/guide.pdf

Niefeld, M. R., Braunstein, J. B., Wu, A. W., Saudek, C. D., Weller, W. E., & Anderson, G. F. (2003). Preventable hospitalization among elderly Medicare beneficiaries with type 2 diabetes. *Diabetes Care, 26*, 1344-1349.

Oster, A., & Bindman, A. B. (2003). Emergency department visits for ambulatory care sensitive conditions: Insights into preventable hospitalizations. *Medical Care, 41*, 198-207.

Ozminkowski, R. J., Goetzel, R. Z., Shechter, D., Stapleton, D. C., Baser, O., & Lapin, P. (2006). Predictors of preventive service use among Medicare beneficiaries. *Health Care Financing Review, 27*(3), 5-23.

Rathore, S. S., Wang, Y., Druss, B. G., Masoudi, F. A., & Krumholz, H. M. (2008). Mental disorders, quality of care, and outcomes among older patients hospitalized

with heart failure: An analysis of the national heart failure project. *Archives of General Psychiatry, 65*, 1402-1408.

Rice, C., & Duncan, D. F. (1995). Alcohol use and reported physician visits in older adults. *Preventive Medicine, 24*, 229-234.

Rice, D. P., Conell, C., Weisner, C., Hunkeler, E. M., Fireman, B., & Hu, T. W. (2000). Alcohol drinking patterns and medical care use in an HMO setting. *Journal of Behavioral Health Services & Research, 27*, 3-16.

Saitz, R. (2003). Medical and surgical complications of addiction. In A. S. T. Graham, M. Mayo-Smith, R. Ries, & B. B. Wilford (Eds.), *Principles of addiction medicine* (3rd ed., pp. 1027-1052). Chevy Chase, MD: American Society of Addiction Medicine.

Shelkey, M., & Wallace, M. (1999). Katz Index of independence in activities of daily living. *Journal of Gerontological Nursing, 25*, 8-9.

STATA Corporation. (n.d.). STATA version 9.0 [Computer software]. College Station, TX: Author.

Wang, M. C., Rosen, A. K., Kazis L., Loveland, S., Anderson, J., & Berlowitz, D. (2000). Correlation of risk adjustment measures based on diagnoses and patient self-reported health status. *Health Services & Outcomes Research Methodology, 1*, 351-365.

Weissman, J. S., Gatsonis, C., & Epstein, A. M. (1992). Rates of avoidable hospitalization by insurance status in Massachusetts and Maryland. *JAMA, 268*, 2388-2394.

Wooldridge, J. M. (2006). *Introductory econometrics* (3rd ed.). Mason, OH: Thomson Southwestern.

Zeng, F., O'Leary, J. F., Sloss, E. M., Lopez, M. S., Dhanani, N., & Melnick, G. (2006). The effect of Medicare health maintenance organizations on hospitalization rates for ambulatory care-sensitive conditions. *Medical Care, 44*, 900-907.

Zhao, Y., Ash, A. S., Ellis, R. P., & Slaughter, J. P. (2002). Disease burden profiles: An emerging tool for managing managed care. *Health Care Management Science, 5*, 211-219.

Zhao, Y., Ellis, R. P., Ash, A. S., Calabrese, D., Ayanian, J. Z., Slaughter, J. P., et al. (2001). Measuring population health risks using inpatient diagnoses and outpatient pharmacy data. *Health Services Research, 36*, 180-193.

A randomized-controlled trial of computerized alerts to reduce unapproved medication abbreviation use

Jennifer S Myers,[1,2] Sattar Gojraty,[1] Wei Yang,[3] Amy Linsky,[4] Subha Airan-Javia,[1] Rosemary C Polomano[5]

[1]Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA
[2]Department of Clinical Effectiveness and Quality Improvement, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA
[3]Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA
[4]Boston University School of Public Health, Boston, Massachusetts, USA
[5]University of Pennsylvania School of Nursing, Philadelphia, Pennsylvania, USA

Correspondence to
Dr Jennifer S Myers, Department of Medicine, Hospital of the University of Pennsylvania, 3400 Spruce Street, Penn Tower Suite 2009, Philadelphia, PA 19104, USA; jennifer.myers@uphs.upenn.edu

A portion of this work was presented at the Society of Hospital Medicine National Meeting in San Diego, California in May 2008 and the National Patient Safety Foundation Meeting in Nashville, Tennessee in May 2008.

Received 15 December 2009
Accepted 3 November 2010
Published Online First
3 December 2010

ABSTRACT

Abbreviation use is a preventable cause of medication errors. The objective of this study was to test whether computerized alerts designed to reduce medication abbreviations and embedded within an electronic progress note program could reduce these abbreviations in the non-computer-assisted handwritten notes of physicians. Fifty-nine physicians were randomized to one of three groups: a forced correction alert group; an auto-correction alert group; or a group that received no alerts. Over time, physicians in all groups significantly reduced their use of these abbreviations in their handwritten notes. Physicians exposed to the forced correction alert showed the greatest reductions in use when compared to controls (p=0.02) and the auto-correction alert group (p=0.0005). Knowledge of unapproved abbreviations was measured before and after the intervention and did not improve (p=0.81). This work demonstrates the effects that alert systems can have on physician behavior in a non-computerized environment and in the absence of knowledge.

INTRODUCTION

Medication errors are responsible for a large number of adverse drug events in patients each year, and the use of medication abbreviations accounts for a subset of these errors.[1–3] For years, professional organizations and regulatory agencies have emphasized the danger of medication abbreviations and have mandated the elimination of the most error-prone abbreviations in medical documentation.[4–7] Because the majority of abbreviation errors originate during medication prescribing,[8] strategies to reduce abbreviations have largely focused on education to modify physician documentation.[9–11] Promulgation of a 'Do Not Use' list of abbreviations created by the Institute of Safe Medication Practices, included in the National Patient Safety Goals, and endorsed by the Joint Commission[4] has served as the primary educational campaign, but there is poor compliance among hospital staff with this practice.[12]

From 2004 to 2006, 643 151 medication errors were reported to the United States Pharmacopeia MEDMARX program from 628 facilities, and 29 974 (4.7%) of these errors involved abbreviation use.[8] Eighty-one per cent of the abbreviation errors occurred during medication prescribing, and 0.3% of errors resulted in patient harm. While a direct association between abbreviations and medication errors has been established, little is known about the best ways to eliminate or reduce abbreviation use.

Medication errors, and in some settings adverse drug events, have been reduced with the adoption of computerized provider order entry (CPOE) and clinical decision support systems (CDSS).[13 14] However, despite widespread acceptance of the benefits of health information technology and national agendas to expand their use,[15 16] in 2008 only 17% of US hospitals had adopted CPOE.[17] As a result, opportunities to introduce medication abbreviations into handwritten documentation remain a source of medication errors and patient harm.

Although a direct link between abbreviations in handwritten notes and medication prescribing errors has not been established, written documentation in the form of handwritten notes and electronic entries with free text is capable of introducing abbreviations that can be misinterpreted and cause errors.[18 19] As the integration of electronic medical records expands nationally, it is important to understand how computerized alerts and clinical decision support influence the knowledge and behaviors of healthcare professionals. Given the paucity of research around electronic interventions to decrease unsafe medication abbreviation use, we conducted a randomized-controlled trial to evaluate the effects of computerized alerts designed to reduce unapproved abbreviations on the frequency of use of these abbreviations in an electronic progress note system and in the non-computer-assisted handwritten documentation of physicians.

METHODS

Study design overview

This study was conducted between July 2006 and June 2007. All internal medicine interns (N=59) at the Hospital of the University of Pennsylvania enrolled in the study at the beginning of their internship. The University of Pennsylvania Institutional Review Board approved the study and granted a waiver of written informed consent. As a condition of the Institutional Review Board approval, participating interns were told that they were part of an ongoing study to examine the effects of computerized interventions designed to reduce unapproved abbreviations but given minimal information about the study. Specific details of the study were withheld to avoid biasing the results. No sources of external funding supported this investigation.

Overview of information systems and medical records at the study site

The hospital has a CPOE system for physician orders and diagnostic test results. The inpatient

medical record is a hybrid of electronic and handwritten documentation. At the time of this study, all history and physical exams (H & Ps) were handwritten, and daily progress notes were created using a customized electronic progress note template. These templates were created within a data-storage program at the University of Pennsylvania (Medview, Microsoft ASP.NET v1.1). All clinical data including medications was entered into the computer by hand and copied forward for daily editing. Progress notes were printed daily and placed into the paper medical record where attending physicians could review and addend them by hand.

Design of the clinical decision support system

The authors of this study designed the clinical decision-support system. Alerts were integrated into the customized electronic progress note templates. The progress note application was modified with regular expression pattern-matching code on the client and server to recognize abbreviations from the Joint Commissions' 'Do Not Use' abbreviation list[4] anywhere in the text and medication lists of the notes, and to generate an alert based on the participant's study group assignment. The application tracked the number of alerts generated for each note. The 'Do Not Use' abbreviation list includes: QD, QOD, MS04, MgSO4, U, IU, trailing zeros, and naked decimal points (table 1). The abbreviation 'MS—morphine sulfate' was not included in our study because we believed that it would reduce the specificity of the alert system. 'MS' within medical record documentation is commonly used to denote terms other than morphine sulfate such as mental status, mitral stenosis, or multiple sclerosis. Since we could not isolate the alert to the medication list, we believed that including it would cause alerts for the non-medication 'MS' terms and lead to documentation errors and clinician frustration.

Randomization

Fifty-nine interns were randomized to one of three study arms using a computer-generated random numbers table. Group 1 received a forced or 'hard-stop' alert that appeared when interns attempted to enter unapproved abbreviations into the electronic progress notes. This alert identified the unapproved abbreviation(s), informed interns of the correct non-abbreviated notation, and forced them to correct the abbreviation before allowing them to save or print their note (figure 1). Group 2 also received an alert when an unapproved abbreviation was entered, but instead of forcing the interns to make a correction, an autocorrection feature displayed the correction and automatically replaced the abbreviation with the acceptable non-abbreviated notation (figure 1). Group 3 was a control group and received no alerts. The alert intervention was introduced 3 months after the study began to allow for observation of baseline medical record documentation practices (figure 2).

Participants did not receive any training sessions about the computerized enhancements and were not informed of their study-group assignment. All groups were exposed to the hospital's standard education for unapproved abbreviations that consisted of reminders to avoid unapproved abbreviations on printed medical note templates.

Table 1 Official 'Do Not Use' List of Abbreviations from the Joint Commission[4]

Do not use*	Potential problems	Use instead
U (unit)	Mistaken for '0' (zero), the number '4' (four) or 'cc'	Write 'unit'
IU (international unit)	Mistaken for IV (intravenous) or the number 10 (ten)	Write 'International Unit'
Q.D, QD, q.d, qd (daily)	Mistaken for each other	Write 'daily'
Q.O.D,. QOD, q.o.d., qod (every other day)	Period after the Q mistaken for 'I' and the 'O' mistaken for 'I'	Write 'every other day'
Trailing zero (X.0 mg)	Decimal point is missed	Write X mg
Lack of leading zero (.X mg)		Write 0.X mg
MS	Can mean morphine sulfate or magnesium sulfate; confused for one another	Write 'morphine sulfate'
MSO4 and MgSO4		Write 'magnesium sulfate'

*Applies to all orders and all medication-related documentation that is handwritten (including free-text computer entry) or on preprinted forms.

Primary outcomes

Retrospective reviews of the medication lists within interns' non-computer-assisted handwritten H & Ps were performed at study conclusion. The medication lists were reviewed to identify the presence or absence of the seven previously defined unapproved abbreviations, and an audit tool was developed to measure the frequency of these abbreviations. In order to estimate the opportunity for an abbreviation error, we had to define the frequency of an absence of the abbreviation. This absence was defined as the frequency with which a correct notation (non-abbreviation) was used. The total opportunity for error was the sum of all present and absent abbreviations. The percentage of unapproved medication abbreviations was defined as the number of abbreviation errors divided by the opportunity for error. Four study investigators (SG, JM, AL, SA) independently reviewed 100 H & Ps to assess reliability of the audit tool. One study investigator (SG) reviewed the remaining H & Ps after reliability statistics were obtained. All reviewers were blinded to the participants' study-group assignment.

A maximum of 15 H & Ps were randomly selected for each participant during each of four study time periods to determine the rate of unapproved abbreviations used over time (figure 2). The numbers of available H & Ps per quarter varied because interns were on vacation, on outpatient rotations, or on rotations at affiliated hospitals. If an intern did not have 15 H & Ps available during a study period, the total number of available H & Ps for that time period was used in the analysis. Interns spent an average of 7 months on inpatient rotations at the hospital where the study was performed.

Secondary outcomes

Secondary outcomes included the frequency of computerized alerts over time and intern knowledge of unapproved abbreviations before and after the study intervention. Knowledge was measured by a test created by the investigators in which interns were asked to identify error-prone abbreviations (unapproved) versus acceptable abbreviations (approved) out of a list of 30 total abbreviations in random order. Additional test items surveyed interns about prior exposure to medication safety education, experiences during medical school (pre-test), and their attitudes about the alerts (post-test).

Statistical analysis

Baseline characteristics among the three groups were compared using the Fisher exact test for categorical variables and Kruskal–Wallis test for continuous variables. Comparisons of the percentages of unapproved medication abbreviations at follow-up periods were done by fitting a pooled logistic regression model which included group indicator, indicator of follow-up time, and their interaction terms (group×follow-up time) as predictors. In this model, each H & P was considered a separate

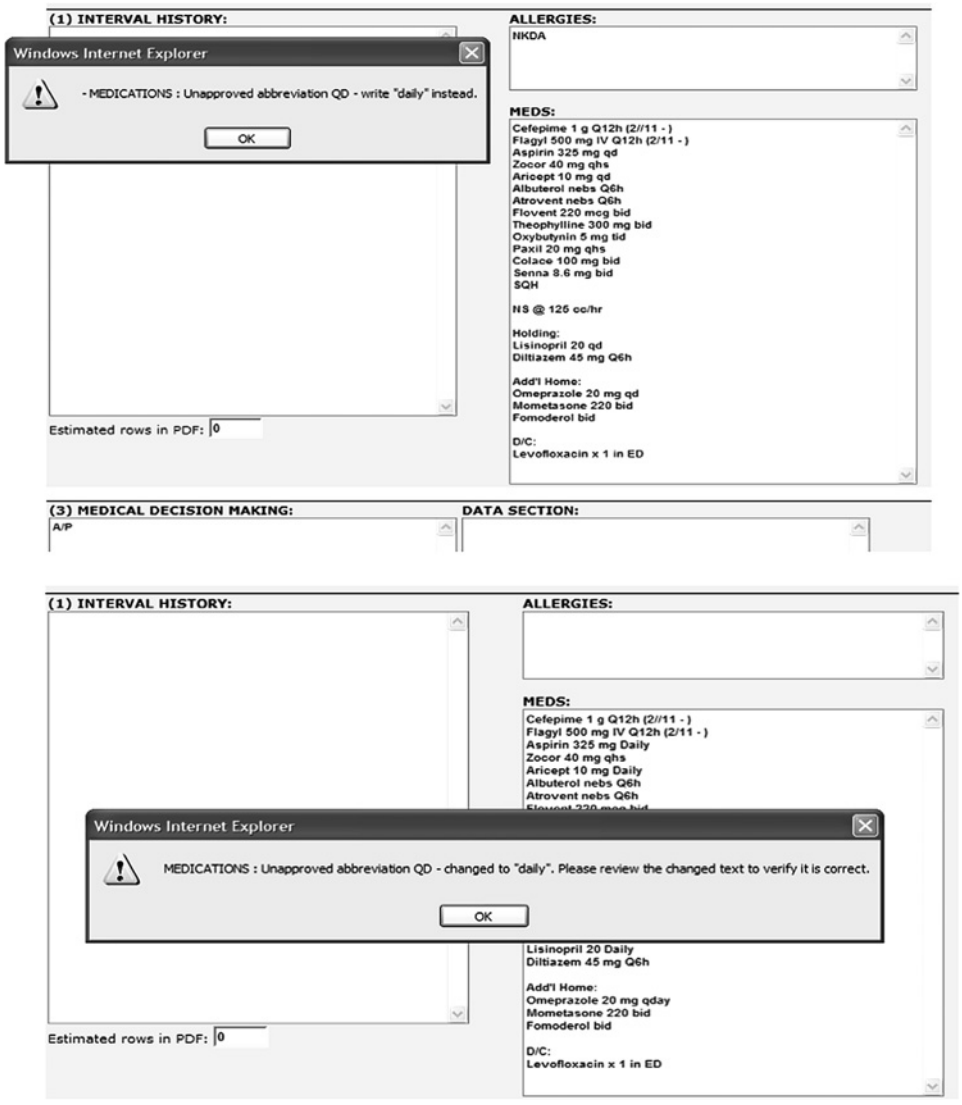


Figure 1 Examples of computerized alert screens used in the intervention. (Top) Example of alert with forced functionality ('hard stop'). (Bottom) Example of alert with an auto-correction feature.

record. Compared to a method in which the percentages for each subject were calculated first and then compared across groups, this method may result in a better precision of estimates by putting less weight on subjects who had fewer H & Ps. Robust variance estimation with a first-order autoregressive (AR (1)) working correlation structure was used to account for repeated measurements within each subject. Both estimated percentages for each group at each follow-up period and the p values for comparisons of the estimated percentages between groups and their change within each group were reported. κ Statistics were used to assess the degree of congruency among four raters of the medication list audit tool. Pre- and post-knowledge differences between the groups were assessed using the Wilcoxon signed-rank test and overall with the Kruskal–Wallis test. All analyses were carried out in SAS version 9.1.

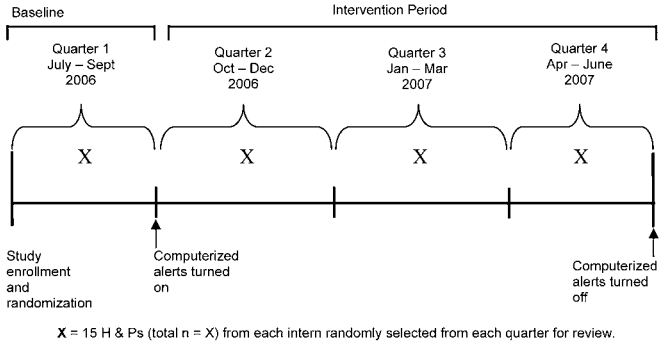


Figure 2 Study design overview.

RESULTS

One hundred per cent (n=59) of interns randomized completed the study and had primary data available for review. Interns had previously attended 23 different medical schools, and their characteristics are listed in table 2. There was no difference among the three groups in their ability to correctly identify unapproved medication abbreviations at baseline (p=0.20).

The median number of H & Ps per study period was 12 (range 0–39). Of the 236 study periods available (59 interns×4 study periods each), there were 13 interns (four control, four hard stop, and five auto-correct) who had one study period with zero H & Ps to review. Based on these numbers, a total of 2371 H & Ps were evaluated with a mean of 42 H & Ps per intern (median=41, range 20–59).

Overall there were 4191 total opportunities to use a 'Do Not Use' abbreviation. Unapproved abbreviations were used 1832 times or 44% of the time. The median number of abbreviation

Table 2 Characteristics and baseline knowledge of unapproved medication abbreviations among study participants (interns)

	Control group (N=19)	Hard stop alert group (N=20)	Auto-correction alert group (N=20)	Total (N=59)	p Value
Men, n (%)	9 (47%)	7 (35%)	9 (45%)	25 (42%)	0.80
Received education in medical school about medication errors related to abbreviations, n (%)	11 (58%)	15 (75%)	9 (45%)	35 (59%)	0.15
Involved in the care of a patient who experienced a medication error, n (%)	10 (53%)	13 (65%)	13 (65%)	36 (61%)	0.67
Baseline knowledge of error-prone abbreviations, mean (IQR)*	8.0 (3.46 to 11)	8.7 (1.87 to 10)	9.2 (2.78 to 11)	8.6 (2.77 to 11)	0.20

*Number of unapproved abbreviations identified correctly out of a list of 11.
IQR, interquartile range.

errors per H & P was 2.5 (range 0–17). The frequency of errors for each abbreviation type was as follows: QD 1672 (91.4%); U 92 (5%); QOD 39 (2.1%); naked decimal point: 20 (1.1%); trailing zero: 5 (0.3%); MgSO4 1 (0.06%); MSO4 1 (0.06%); IU 0 (not written in any H & Ps). Many H & Ps contained more than one abbreviation error. The inter-rater agreement for the medication list audit tool was excellent (κ=1).

Primary outcome

The percentages of unapproved medication abbreviations in each quarter are shown in table 3. At baseline (Quarter 1), there were no significant differences in the frequency of unapproved abbreviations in the non-computer-assisted handwritten notes among the three study groups (p=0.54) (table 3). Interns in each group significantly reduced their use of non-computer-assisted, written unapproved abbreviations over time (control: p=0.004; hard stop: p<0.0001; autocorrect p=0.04) (figure 3). When compared with controls, interns in the hard-stop alert group had a lower rate of unapproved abbreviations in their non-computer-assisted handwritten notes during the alert intervention (p=0.02), whereas interns in the auto-correction group did not (p=0.21). Interns in the hard-stop alert group had a significantly lower rate of unapproved abbreviations in their non-computer-assisted handwritten notes when compared with interns in the auto-correction group. (p=0.0005).

Secondary outcome

The number of alerts that fired decreased over time (p<0.01) in both alert intervention groups. There was a trend toward fewer alerts firing in the 'hard-stop' group compared with the auto-correction group (p=0.06).

Forty-seven interns (80%) completed the knowledge test at study conclusion. Knowledge of unapproved abbreviations did not improve after the alert intervention. At baseline, interns correctly identified 8.6 (range 5–11; 73% correct) unapproved abbreviations compared with 9.0 (range 5–11; 82% correct) after the intervention. This was true for the entire sample (p=0.81), within individual groups (hard-stop, p=0.67; auto-correction, p=0.09; control, p=0.31), and between groups (p=0.39). Intern attitudes about the alerts were assessed in the post-test. Among interns who received alerts, nine (26%) believed that the alerts interfered with their ability to efficiently complete their documentation, 14 (41%) did not, and 11 (32%) were neutral. No attitude differences were detected between the two alert groups.

Power analysis

Repeated measurements in this study allowed for an increase in statistical power to detect treatment differences among study conditions. The correlation between two successive measurements was 0.62. In randomized trials with repeated measurements, an important determinant of the minimum detectable difference is the design effect, which is defined as 1+ICC(k−1), where ICC is the intraclass correlation coefficient, that is, the correlation between two successive measurements, and k is the number of repeated measurements. The three repeated measurements in this study yielded a design effect of 2.24. The effective sample size for each group is 20×3/2.24=27. Using an estimated SD of the percentage of unapproved abbreviations of 0.5 (table 3), the study has 80% power to detect a difference in percentages of abbreviations of 0.39 between groups with a type I error of 0.05.

DISCUSSION

This randomized-controlled trial compared two types of computerized alerts designed to reduce the use of unapproved medication abbreviations by physicians. We demonstrated that alerts embedded within an electronic progress note program reduced the use of abbreviations within the electronic program (as measured by frequency of alerts fired) and within the non-computer-assisted handwritten H & Ps authored by physicians over the same time period. Alerts with a forced correction feature decreased the use of abbreviations to a much greater extent than alerts with an auto-correction feature. Moreover, an unanticipated but particularly interesting finding in our study was that reductions in abbreviation use were observed in a control group who were unexposed to alerts, but who were exposed to the overall study environment.

Table 3 Summary of the percentage of unapproved medication abbreviations in handwritten notes among interns exposed to no alerts, hard stop alerts, or auto-correction alerts in an electronic note writing program

	Control group			Hard stop alert group			Auto-correction alert group		
Follow-up time	Total no of opportunities for error	No of unapproved medication abbreviations	Percentage of unapproved medication abbreviations	Total no of opportunities for error	No of unapproved medication abbreviations	Percentage of unapproved medication abbreviations	Total no of opportunities for error	No of unapproved medication abbreviations	Percentage of unapproved medication abbreviations
Quarter 1 (baseline)	317	191	0.60	426	214	0.50	231	155	0.67
Quarter 2	386	188	0.49	299	79	0.26	182	99	0.54
Quarter 3	281	63	0.22	373	31	0.08	252	150	0.60
Quarter 4	366	49	0.13	324	36	0.11	271	94	0.35

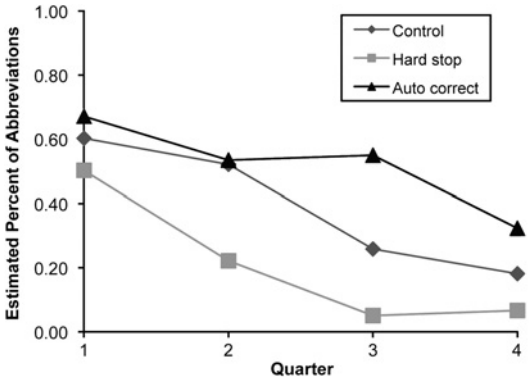


Figure 3 Comparisons of estimated percentages of unapproved abbreviations in handwritten notes across different groups. Error rate for Quarter 1 (baseline) was estimated using the raw data; error rates for Quarters 2–4 were estimated using a pooled logistic regression model which includes group indicator, indicator of follow-up time, and their interaction terms (group×follow-up time), and specifies autoregression working correlation matrix. p Value for comparisons of error rates across the three groups at baseline is 0.54. p Values for trend test within each group are 0.004, <0.0001, and 0.04 for control, hard stop, and auto-correction group respectively. p Value for comparisons of the error rate between hard stop and control groups is 0.02; 0.21 between auto-correction and control groups, and p=0.0005 between auto-correction and hard stop groups.

Eliminating error-prone medication abbreviations has been extremely challenging for hospitals, and there are very few effective interventions in the literature for this vexing problem. An educational intervention designed to reduce prescribing errors in the handwritten medication orders of residents reduced overall prescribing errors among surgery but not medicine residents.[11] Enforcement strategies at the level of medical staff leadership proved more effective than education alone in a single study[10]; however, enforcing physician accountability for documentation skills is difficult. Given the strong and repeated association between abbreviation use and medication errors,[3 8] it will be necessary and important for healthcare leaders to use multiple strategies to improve this unsafe and therefore unacceptable practice. Health information technology is just one of those strategies. As demonstrated in this study, a clinical decision-support system designed to reduce abbreviations may be an effective addition to administrative oversight and routine education.

Of the 2371 H & Ps reviewed, there were 4191 unapproved abbreviations noted, which equates to approximately two unapproved abbreviations per H & P. On the surface, this average seems low considering the high numbers of patients in our hospital treated with multiple medications. However, when considering the frequency of these occurrences in H & Ps (range 0–17; median=2.5), one can see that significant abbreviation use with the opportunity for medication errors exists. Significant reductions in abbreviations were demonstrated in the non-computer-assisted handwritten notes over time and across all three study groups, further reducing the abbreviation errors in the H & Ps.

The ability for health information technology to intercept unsafe practices and prevent serious medication errors has been described.[20–22] Improvements in medication safety with the use of CDSS occur through both direct and indirect effects. Direct effects alter medication prescribing or management at the time practitioners interact with system. Indirect or 'spillover' effects result from the carry-over into practice of knowledge or behaviors learned during exposure to the system.[23] Few studies of CDSSs have been designed to measure indirect effects. Glassman *et al* reported that exposure to automated drug alerts had little effect on the recognition of selected drug–drug or drug–condition interactions as measured by a cross-sectional survey.[23] Studies of drug-utilization reviews describe indirect effects of interventions on future clinician behavior.[24 25] For example, a time-series study that involved mailing letters to physicians about drug interactions and monitoring their subsequent prescribing patterns found no effect on future prescribing behavior as a result of the intervention.[24] In contrast, our study found large indirect effects by demonstrating significant reductions in the frequency of medication abbreviations in physicians' non-computer-assisted handwritten notes when they were prompted to correct the abbreviations in the electronic notes over the same time period.

The rapid expansion of alert systems in medical informatics calls for more research comparing the effects of different alert systems on the same outcome. Previous studies have described the over-riding of drug safety alerts[26] and demonstrated that the nature of alerts influences clinician behavior.[22 27 28] Thus, the inability to detect significant reductions in abbreviation use in the non-computer-assisted handwritten notes in the auto-correction alert group compared with the forced correction alert group is not surprising. One reason for this is that interns who received auto-correction alerts disregarded the educational message or simply acknowledged the alert without reading the information given human factors such as time pressure, competing priorities, and alert fatigue. In contrast, interns exposed to the forced alert were unable to complete their electronic notes without making manual corrections. It is known that mere repetition facilitates long-term memory,[29 30] and it may be that by forcing physicians to correct abbreviations, their knowledge of these abbreviations was solidified and translated into improvements in written practice. In summary, our study found direct evidence that passive alerts do little to influence clinician behavior. Additional studies will be important to substantiate these findings and advance the field of health informatics.

Reductions in abbreviation use in the control group were not anticipated by the investigators, but there are several possible explanations for the observation. Experimental diffusion, which occurs when a treatment effect applied to one group unintentionally spills over and contaminates another group,[31] may explain our findings. Interns in the control group were working in a study environment designed to modify physician behavior. Even though they were not directly exposed to alerts, their behavior may have been influenced by the improving documentation patterns of the interns exposed to the intervention who worked with them. Diffusion of effects threatens the internal validity of research, but it is difficult to control for in quality improvement research. While it is possible that the hospitals' educational strategies to reduce unapproved abbreviations contributed to the documentation improvements, this seems unlikely given the historical failure of routine education related to abbreviation avoidance.[10–12]

Despite the improvements in documentation practices, we failed to find any significant improvements in physician knowledge of unapproved abbreviations. This apparent 'disconnect' in knowledge versus practice is intriguing and has been demonstrated previously by Glassman *et al*.[23] There are several possible explanations for this finding. Our sample size may have been too small to detect a meaningful difference. The participants had varying degrees of exposure to the abbreviation

alerts and thus may have been unable to remember the abbreviations when they were presented to them in the post-test. The alerts may not have been perceived as important or relevant to the interns, especially since they were alerted when writing notes rather than when ordering medications. Nonetheless, it is hard to ignore the substantial reductions in abbreviation use in the non-computer-assisted handwritten notes as a result of the intervention, and this could be interpreted as a surrogate for knowledge acquisition.

The Joint Commission has strictly prohibited the use of seven common and unsafe medication abbreviations,[4] and The Institute for Safe Medication Practices has promulgated a list of over 50 abbreviations that have been associated with harm in their error-reporting systems and should never be used.[5] However, given the fact that any medication abbreviation creates an opportunity for misinterpretation and error potential, some organizations have attempted to limit all medication abbreviations by creating policies with 'approved' (rather then unapproved) abbreviation lists in which all medication abbreviations are prohibited. Since clinicians are in the habit of using medication abbreviations frequently, it is unlikely that one intervention alone will eliminate this practice, and it will be necessary to consider electronic interventions such as this to curb their use in free text entries in prescription writing and medical records.

Our study has several limitations. Because our hospital has an integrated CPOE system, we were unable to assess whether our intervention would have affected handwritten medication prescribing errors related to abbreviations. Handwritten abbreviations in prescriptions present a larger risk to patients than handwritten abbreviations in medical records. However, it is possible that documentation skills learned by physicians in an electronic environment and practiced in handwritten notes will carry over into their future handwritten or electronic free text prescriptions, and recommendations for prescriptions are often made in medical record documentation, so the potential for abbreviation errors exists even outside of the prescription-writing environment. Feasibility issues prevented retrospective reviews of the handwritten medication prescriptions of study participants.

We did not study the documentation practices of the participants after the alerts were turned off. Consequently, we cannot be certain of the long-term sustainability of our intervention and whether the documentation improvements would have improved further, plateaued, or waned had the alerts been turned off or continued. Additionally, given that exposure to the alerts was not continuous over time based upon the sequence of intern rotations and that these alerts varied in frequency among all interns, we may have reduced our ability to detect important differences among the groups and within certain participants. We did not evaluate documentation practices in the year(s) prior to our intervention and thus cannot completely exclude the possibility that a trend towards reduction in unapproved abbreviations occurred from a natural history effect encountered with introducing electronic platforms for documentation.

Finally, our study has several features that may limit its generalizablity. We studied only interns at a single academic medical center with a hybrid information system comprising both paper and electronic documentation. Since many organizations currently practice in hybrid systems, and many physicians practice in multiple information systems over the course of their career, we believe that the information related to the secular trends in physician non-computer assisted handwritten notes as a result of exposure to computerized alerts is relevant. Compared with interns, residents, attending physicians, non-physician providers, or practitioners in community hospitals may have responded differently to the intervention; however, there are elements of practitioner performance that are not unique to interns or academic medical centers, and some generalizations can be made from this study. The undergraduate and graduate medical training years are an ideal time to introduce information technology designed to improve medication safety, since trainees have not yet been influenced by unsafe medication documentation practices in the hospital and may be more open to changes in practice.

In summary, our study contributes important information to the health information technology literature by describing the effect that CDSS can have on physician behavior in the absence of knowledge and demonstrating that an informatics intervention can create large behavioral changes in a control group unexposed to the actual intervention but exposed to the study environment in which the intervention was performed. The methods used in this study to examine the indirect effects of health information systems to modify physician behavior outside of the electronic environment are unique and may have relevance for other health information technology interventions. We have established a methodology within a randomized-controlled trial to evaluate the effects of alerts embedded within a clinical decision-support system on physician knowledge and practice. Estimates for the percentage of unapproved abbreviation use were calculated based on the number of opportunities for error and offer additional endpoints to measure practice changes with technology-based interventions. These estimates can be used to determine samples sizes for adequate statistical power to evaluate the effects of interventions to reduce medication errors and test information systems in patient safety research. We found that alerts for unapproved medication abbreviations within electronic medical record systems are effective in changing physician documentation and thus promoting medication safety. Given that many healthcare organizations do not have fully integrated health information technology systems, researchers and patient safety leaders will continue to be challenged with ways to promote safe medication practices through electronic tools and education.

Acknowledgments The authors are grateful to L Fleisher and S Williams, for their comments on earlier versions of this manuscript. We also thank S Kratowicz for his technical support in building the computerized alerts.

Competing interests None.

Ethics approval Ethics approval was provided by the University of Pennsylvania IRB.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Kohn LT,** Corrigan JM, Donaldson MS. *To Err is Human: Building a Safer Health System*. Washington, DC: (US) Committee on Quality of Health Care in America, Institute of Medicine, National Academy Press, 2000.
2. **Lesar TS,** Briceland L, Stein DS. Factors related to errors in medication prescribing. *JAMA* 1997;**277**:312–17.
3. **Barker KN,** Flynn EA, Pepper GA, *et al*. Medication errors observed in 36 healthcare facilities. *Arch Intern Med* 2002;**162**:1897–903.
4. The Joint Commission Accreditation Program: Hospital National Patient Safety Goals. NPSG 02.02.01 accessed at http://www.jointcommission.org/NR/rdonlyres/31666E86-E7F4-423E-9BE8-F05BD1CB0AA8/0/HAP_NPSG.pdf page 5 on 8 Jun 2009.
5. ISMP and FDA Campaign to eliminate use of error-prone abbreviations. http://www.ismp.org/tools/abbreviations (accessed 8 Jun 2009).
6. **Landers SJ.** Goodbye to U: safety campaign nixes abbreviations. *Am Med News* 17 July 2006;**48**(27). http://www.ama-assn.org/amednews/2006/07/17/hll20717.htm (accessed 8 Jun 2009).
7. **von Eschenbach AC.** Eliminating error-prone notations in medical communications. *Expert Opin Drug Saf* 2007;**6**:233–4.

8. **Brunetti L,** Santell JP, Hicks RW. The impact of abbreviations on patient safety. *Jt Comm J Qual Pt Saf* 2007;**33**:576—83.
9. **Abushaiqa ME,** Zaran FK, Bach DS, *et al*. Educational interventions to reduce use of unsafe abbreviations. *Am J Health Syst Pharm* 2007;**64**:1170—3.
10. **Traynor K.** Enforcement outdoes education at eliminating unsafe abbreviations. *Am J Health Syst Pharm* 2004;**61**:1314, 1317, 1322.
11. **Garbutt J,** Milligan PE, McNaughton C, *et al*. Reducing medication prescribing errors in a teaching hospital. *Jt Comm J Qual Patient Saf* 2008;**34**:528—36.
12. **Anon.** Compliance data for the Joint Commissions' 2004 and 2005 national patient safety goals. *Jt Comm Perspect* 2005;**25**:7—8.
13. **Bates DW,** Gawande AA. Improving safety with information technology. *N Engl J Med* 2003;**25**:2526—34.
14. **Chaundhry B,** Wang J, Wu S, *et al*. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006;**144**:742—52.
15. *National Committee for Vital and Health Statistics, Information for Health: A Strategy for Building the National Health Information Infrastructure*. Washington, DC: US Department of Health and Human Services, 15 Nov 2001.
16. **Thompson T,** Brailer D. *The Decade of Health Information Technology: Delivering Consumer-Centric and Information-Rich Health Care: Framework for Strategic Action*. Washington: U.S. Department of Health and Human Services, 21 Jul 2004.
17. **Jha AK,** DesRoches CM, Campbell EG, *et al*. Use of electronic health records in U.S. hospitals. *N Engl J Med* 2009;**360**:1628—38.
18. **Manzar S,** Nair AK, Govind Pai M, *et al*. Use of abbreviations in daily progress notes. *Arch Dis Child Fetal Neonatal Ed* 2004;**89**:374.
19. **Sheppard JE,** Weidner LC, Zakai S, *et al*. Ambiguous abbreviations: an audit of abbreviations in paediatric note keeping. *Arch Dis Child* 2008;**93**:204—6.
20. **Bates DW,** Leape LL, Cullen DJ, *et al*. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* 1998;**280**:1311—16.
21. **Potts AL,** Barr FE, Gregory DF, *et al*. Computerized physician order entry and medication errors in a pediatric critical care unit. *Pediatrics* 2004;**113**:59—63.
22. **Garg AX,** Adhikari NKJ, McDonald H, *et al*. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. *JAMA* 2005;**293**:1223—38.
23. **Glassman PA,** Belperio P, Simon B, *et al*. Exposure to automated drug alerts over time: effects on clinicians' knowledge and perceptions. *Med Care* 2006;**44**:250—6.
24. **Hennessy S,** Bilker WB, Zhou L, *et al*. Retrospective drug utilization review, prescribing errors, and clinical outcomes. *JAMA* 2003;**290**:1494—9.
25. **Lipton HL,** Bird JA. Drug utilization and review in ambulatory settings: state of the science and directions for outcomes research. *Med Care* 1993;**31**:1069—982.
26. **Van Der Sijs H,** Aarts J, Vulto A, *et al*. Overriding of drug safety alerts in computerized physician order entry. *J Am Med Inform Assoc* 2006;**13**:138—47.
27. **Spina JR,** Glassman PA, Belperio P, *et al*; Primary Care Investigative Group of the VA Los Angeles Healthcare System. Clinical relevance of automated drug alerts from the perspective of medical providers. *Am J Med Qual* 2005;**20**:7—14.
28. **Glassman PA,** Simon B, Belperio P, *et al*. Improving recognition of drug interactions. Benefits and barriers to using automated drug alert. *Med Care* 2002;**40**:1161—71.
29. **Naka N,** Naoi H. The effect of repeated writing on memory. *Mem Cognit* 1995;**23**:201—12.
30. **Dark VJ,** Loftus GR. The role of rehearsal in long-term memory performance. *J Verbal Learn and Verbal Behav* 1976;**15**:479—90.
31. **Rubinson L,** Neutens JJ. *Research Techniques for the Health Sciences*. 3rd edn. San Francisco, CA: Pearson Education Inc, 2002.

A randomized-controlled trial of computerized alerts to reduce unapproved medication abbreviation use

Jennifer S Myers, Sattar Gojraty, Wei Yang, et al.

JAMIA 2011 18: 17-23 originally published online December 3, 2010
doi: 10.1136/jamia.2010.006130

"Gray area" alcohol consumption and the U.S. Dietary Guidelines: A comment on Dawson and Grant (2011)

Dear Editor:

The research by Dawson and Grant (*Journal of Studies on Alcohol and Drugs,* May 2011) exploring the "gray area" of alcohol consumption was a well-executed and important article, demonstrating that drinking at levels initially proposed for the 2010 Dietary Guidelines for Americans was associated with an increased risk of prevalent and incident alcohol dependence, incident alcohol-related interpersonal problems, and prevalent job loss compared with the limits in the 2005 Dietary Guidelines. The proposed revision would have changed the limits from *daily* consumption not exceeding 2 drinks for men and 1 for women to *average* consumption not exceeding 2/1 drinks for men/women, while stipulating that risky drinking precluded drinking 5/4 or more drinks daily (Dawson and Grant, 2011).

Their findings underscore the wisdom of the ultimate decision to retain the 2005 Dietary Guidelines for Americans limits in the 2010 version (U.S. Department of Health and Human Services and U.S. Department of Agriculture, 2005, 2010) and illustrates that the mere absence of clearly risky drinking does not necessarily constitute low-risk drinking, particularly in a document meant to apply to the entire adult population of the United States. Because most people drinking within 2/1 average limit drink 10 or fewer days per month (Naimi et al., 2010), the effect of the proposed change would have been to condone drinking up to 4 drinks daily for men or up to 3 drinks daily for women, as long as the average limits were not exceeded. Such an interpretation would be analogous to suggesting that driving a car with a blood alcohol concentration of .079% (a level just short of legal intoxication) is safe or desirable. Finally, were a low-risk drinking guideline developed on the basis of average alcohol consumption, the nadir of risk (or the zenith of potential benefit, depending on one's perspective) for all-cause mortality is approximately one third of a U.S. standard drink for women and approximately one half of a drink for men (DiCastelnuovo et al., 2006), both substantially less than what was proposed.

The Dietary Guidelines do not recommend alcohol consumption but rather recommend low-risk drinking limits *for those who choose to consume alcohol*. To date, there have been no randomized trials of low-dose alcohol and any morbidity and mortality outcome, and existing observational studies may be biased in favor of moderate drinkers (Fillmore et al., 2006; Naimi et al., 2005) Both the 2005 and 2010 Dietary Guidelines explicitly recommend against initiating drinking or drinking more frequently on the basis of potential health benefits (U.S. Department of Health and Human Services and U.S. Department of Agriculture, 2005, 2010).

Timothy S. Naimi, M.D., M.P.H.
Section of General Internal Medicine,
Boston Medical Center, and
Associate Professor,
Boston University Schools of Medicine
and Public Health
Boston, MA

References

Dawson, D. A., & Grant, B. F. (2011). The "gray area" of consumption between moderate and risk drinking. *Journal of Studies on Alcohol and Drugs, 72,* 453–458.

Di Castelnuovo, A., Costanzo, S., Bagnardi, V., Donati, M. B., Iacoviello, L., & de Gaetano, G. (2006). Alcohol dosing and total mortality in men and women: An updated meta-analysis of 34 prospective studies. *Archives of Internal Medicine, 166,* 2437–2445.

Fillmore, K. M., Kerr, W. C., Stockwell, T., Chikritzhs, T., & Bostrom, A. (2006). Moderate alcohol use and reduced mortality risk: Systematic error in prospective studies. *Addiction Research and Theory, 14,* 101–132

Naimi, T. S., Brown, D. W., Brewer, R. D. (2010, June). *Binge alcohol consumption and cardiovascular risk factors among "moderate" drinkers in the U.S.* Paper presented at the 33rd Annual Scientific Meeting of the Research Society on Alcoholism, San Antonio, TX.

Naimi, T. S., Brown, D. W., Brewer, R. D., Giles, W. H., Mensah, G., Serdula, M.K., . . . Stroup, D. F. (2005). Cardiovascular risk factors and confounders among nondrinking and moderate-drinking U.S. Adults. *American Journal of Preventive Medicine, 28,* 369–373.

U.S. Department of Health and Human Services and U.S. Department of Agriculture. (2005). *Dietary Guidelines for Americans, 2005.* 6th Edition, Washington, DC: U.S. Government Printing Office, January 2005. Available at: http://www.health.gov/dietaryguidelines/dga2005/document/pdf/DGA2005.pdf

U.S. Department of Health and Human Services and U.S. Department of Agriculture. (2010). *Dietary Guidelines for Americans, 2010.* 7th Edition, Washington, DC: U.S. Government Printing Office. Available at: http://www.cnpp.usda.gov/Publications/DietaryGuidelines/2010/PolicyDoc/PolicyDoc.pdf

AMERICAN JOURNAL OF
Preventive Medicine

VOLUME 41(5) www.ajpmonline.org NOVEMBER 2011

A Journal of the
ACPM American College of Preventive Medicine & APTR Association for Prevention Teaching and Research

ELSEVIER

The Cost of Alcohol and Its Corresponding Taxes in the U.S.

A Massive Public Subsidy of Excessive Drinking and Alcohol Industries

Timothy S. Naimi, MD, MPH

In an invaluable update using a standard cost-of-illness approach, Bouchery et al.[1] estimated the financial impact of excessive alcohol consumption in the U.S. for 2006, the most recent year for which data were available. The results ($223.5 billion or $746 per capita that year) are staggering and exceed the costs of the other leading preventable causes of death in the U.S., including cigarette smoking and physical inactivity. This is partly because excessive alcohol consumption involves many second-hand or external costs (i.e., costs that are incurred by those other than the consumers, sellers, or producers of alcohol) and because many alcohol-related outcomes begin at relatively young ages, which results in large future productivity losses and prolonged or recurrent expenditures related to health care and the criminal justice system.

Although this method is not the only way to estimate the cost of public health problems, it is commonly used and generally accepted. The method is conservative in that it does not include costs related to pain and suffering, which can account for more than half of the costs in studies where this is valued using a willingness-to-pay approach. The study by Bouchery et al.[1] is also conservative principally because of limitations around delineating alcohol-attributable health effects, which has a large influence on health and productivity costs. On the other hand, this approach considers the future cost of lost productivity but not averted economic consumption, which is the equivalent of savings that could offset costs. Finally, this is a study about the costs of excessive drinking, and low-risk alcohol consumption may have some economic benefits. In this light, however, one could argue that the costs from excessive drinking are conservative in that they also represent an unmeasured opportunity cost associated with failing to achieve possible savings from low-risk consumption by, for example, raising alcohol prices to the point where they would mitigate excessive drinking and instead promote low-risk drinking among those who drink.

An important advance in this article is that it takes the cost estimate, which is difficult to grasp, and translates it on a cost-per-drink basis that makes it comprehensible to policymakers, as is, for example, the cost per pack of cigarettes. The cost-per-drink metric also facilitates a comparison with current alcohol taxes, which can be derived on a per-drink basis. This, in turn, highlights perhaps the most important contribution of this study, which is to illustrate the gross disparity that exists between the cost of alcohol consumption and its taxes.

Based on federal tax rates for standard alcohol beverage categories (5% alcohol-by-volume [ABV] beer; 12% ABV wine; and 40% ABV liquor), and after weighting those taxes on the basis of beverage-specific consumption in the U.S. and standard drink size (14 g of ethanol per drink), the average federal tax in the U.S. is approximately 8.5 cents per drink.[2-5] Further, since federal alcohol taxes are based on a fixed amount per volume of alcohol, they continuously erode as a result of inflation. For example, the federal beer tax has declined by 41% in real terms since it was last adjusted in 1991. Historically, alcohol taxes accounted for approximately 40% of federal revenues; they now account for less than 0.5% of revenue.[6]

States typically have substantially lower taxes on alcohol than does the federal government (approximately 5 cents per drink). Moreover, although these state taxes are enacted primarily through volume-based excise taxes (similar to federal taxes), in some cases they include ad valorem taxes (alcohol-specific taxes based on a percentage of the price) or general sales taxes. In Massachusetts, my home state, the weighted average tax per drink is only 2.6 cents,[4] which is levied in the form of volume-based excise taxes. In addition, as in a number of other states, alcohol is not subject to the state's 6.25% sales tax or any ad valorem taxes, meaning that alcohol is taxed far less than items such as durable medical equipment, automobiles, or other general merchandise.

The study's estimated $1.90 cost per drink in 2006 has increased to $2.13 per drink in 2011, after adjustment for

From the Section of General Internal Medicine, Boston Medical Center, Boston, Massachusetts

Address correspondence to: Timothy Naimi, MD, MPH, BMC/SGIM, 801 Massachusetts Ave., Rm. 2079, 15 Boston MA 02118; E-mail: tim.naimi@bmc.org.

0749-3797/$17.00
doi: 10.1016/j.amepre.2011.08.001

inflation. Because federal plus state taxes are (and were) approximately 14 cents per drink, this is a current disparity between cost and tax of approximately $2 per drink. In addition, this study found that 41.5% of costs were incurred by drinkers themselves, meaning that 58.5% of costs (or approximately $1.25 per drink in 2011 and $1.11 in 2006) were external to drinkers, a disparity of $1.11 per drink for external costs in 2011 and $0.97 per drink in 2006. Therefore, in 2006 alone, the 117.4 billion standard drinks consumed in the U.S. resulted in $113.9 billion in net (i.e., un-recouped) external costs that accrued to the general public; for comparison, this was almost half the size of the federal budget deficit that year.

External economic costs are an important justification for, and basis of, taxation on items such as alcohol, cigarettes, or businesses that generate environmental pollution.[6] Failing to recoup these external costs amounts to a massive public subsidy, in which the 80% of the U.S. population that doesn't drink or that drinks in low-risk ways pays for costs incurred by excessive drinkers and those who produce, distribute, or sell alcohol. Were this subsidy addressed through commensurate taxation, it would have an adverse effect on alcohol-related businesses. However, by making people pay the true cost of alcohol, the additional revenue would promote economic efficiency through re-allocation to other sectors of the economy and could be used to finance state and federal debt obligations, pay for education or publicly financed health care, or increase personal income by offsetting income or property taxes. Based on the above, increased taxes would result in net savings for most taxpayers, and excessive drinkers would pay almost five times as much per capita as low-risk drinkers according to their differences in consumption. And unlike tobacco taxes, increased alcohol taxes would be borne principally by those who are relatively socially and economically advantaged.[7]

In addition to recouping external economic costs, there is another overriding rationale for increasing alcohol taxes: to improve public health and well-being. Maintaining higher prices generally, and raising alcohol taxes specifically, is the most effective population-based means of preventing and reducing excessive alcohol consumption and related harms.[8] Unlike cost-of-illness valuations, pain and suffering do matter when it comes to health and well-being, and most would value a life more highly than the sum of wages earned while living it. In other words, raising alcohol taxes by an order of magnitude is ultimately a matter of economic fairness that will result in societal benefits that are, to many, beyond measure.

The author wishes to acknowledge Jason Blanchette, MPH, for assistance with the preparation of this manuscript. This work was partly supported by NIH grant AA018377.

The research presented in this paper is that of the authors and does not reflect the official policy of the NIH.

No financial disclosures were reported by the author of this paper.

References

1. Bouchery E, Harwood H, Sacks J, Simon CJ, Brewer RD. Economic costs of excessive alcohol consumption in the U.S., 2006. Am J Prev Med 2011;41(5):516–524.
2. Tax Policy Center. Tax facts. Alcohol tax revenue. State and local alcohol beverage tax revenue, selected years, 1997–2007. www.taxpolicycenter.org/taxfacts/displayafact.cfm?Docid=399.
3. Tax Policy Center. Federal excise taxes reported to or collected by the Internal Revenue Service, Alcohol and Tobacco Tax and Trade Bureau, and Customs Service, by type of excise tax, fiscal years 1996-2008. www.taxpolicycenter.org/taxfacts/Content/PDF/excise_type.pdf.
4. National Institutes of Alcohol Abuse and Alcoholism. Per capita ethanol consumption for states, census regions, and the United States, 1970–2007. (Gallons of ethanol, based on population age 14 and older). www.niaaa.nih.gov/Resources/DatabaseResources/QuickFacts/AlcoholSales/Pages/consum03.aspx.
5. National Institutes of Alcohol Abuse and Alcoholism. What is a standard drink? pubs.niaaa.nih.gov/publications/practitioner/pocketguide/pocket_guide2.htm.
6. Cook P. Paying the tab: the economics of alcohol policy. Princeton NJ: Princeton University Press, 2007.
7. Stahre M, Daley J, Naimi T. The impact of a $.25 per drink tax increase: who really pays the tab? Paper presented at the 3rd North American Congress for Epidemiology; 2011 Jun 21–24; Montreal, Quebec, Canada.
8. Babor T, Caetano R, Casswell S, et al., eds. Alcohol: no ordinary commodity. 2nd ed. New York: Oxford University Press, 2010.

OPEN ACCESS Freely available online

PLoS GENETICS

Genetic Determinants of Serum Testosterone Concentrations in Men

Claes Ohlsson[1,9]*, Henri Wallaschofski[2,9], Kathryn L. Lunetta[3,4,9], Lisette Stolk[5,6,9], John R. B. Perry[7,8,9], Annemarie Koster[9,9], Ann-Kristin Petersen[10,9], Joel Eriksson[1,9], Terho Lehtimäki[11,9], Ilpo T. Huhtaniemi[12,9], Geoffrey L. Hammond[13,9], Marcello Maggio[14], Andrea D. Coviello[3,15], EMAS Study Group[16], Luigi Ferrucci[17], Margit Heier[18], Albert Hofman[6,19], Kate L. Holliday[20], John-Olov Jansson[21], Mika Kähönen[22], David Karasik[23], Magnus K. Karlsson[24], Douglas P. Kiel[3,23], Yongmei Liu[25], Östen Ljunggren[26], Mattias Lorentzon[1], Leo-Pekka Lyytikäinen[11], Thomas Meitinger[27,28], Dan Mellström[1], David Melzer[29], Iva Miljkovic[30], Matthias Nauck[2], Maria Nilsson[1], Brenda Penninx[31], Stephen R. Pye[20], Ramachandran S. Vasan[3,15], Martin Reincke[32], Fernando Rivadeneira[5,6,19], Abdelouahid Tajar[20], Alexander Teumer[33], André G. Uitterlinden[5,6,19], Jagadish Ulloor[15], Jorma Viikari[34], Uwe Völker[33], Henry Völzke[35], H. Erich Wichmann[36,37,38], Tsung-Sheng Wu[13], Wei Vivian Zhuang[4], Elad Ziv[39,40], Frederick C. W. Wu[41¶], Olli Raitakari[42,43¶], Anna Eriksson[1¶], Martin Bidlingmaier[32¶], Tamara B. Harris[9¶], Anna Murray[7,8¶], Frank H. de Jong[5¶], Joanne M. Murabito[3,15¶], Shalender Bhasin[3,15¶], Liesbeth Vandenput[1¶], Robin Haring[2¶]

1 Center for Bone and Arthritis Research, Department of Internal Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, **2** Institute of Clinical Chemistry and Laboratory Medicine, University of Greifswald, Greifswald, Germany, **3** The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, United States of America, **4** Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, **5** Department of Internal Medicine, Erasmus MC Rotterdam, Rotterdam, The Netherlands, **6** Netherlands Consortium of Healthy Ageing, Rotterdam, The Netherlands, **7** Genetics of Complex Traits, Peninsula Medical School, University of Exeter, Exeter, United Kingdom, **8** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **9** Laboratory for Epidemiology, Demography, and Biometry, National Institute on Aging, Bethesda, Maryland, United States of America, **10** Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany, **11** Department of Clinical Chemistry, University of Tampere and Tampere University Hospital, Tampere, Finland, **12** Department of Surgery and Cancer, Hammersmith Campus, Imperial College London, London, United Kingdom, **13** Child and Family Research Institute and Department of Obstetrics and Gynecology, University of British Columbia, Vancouver, Canada, **14** Departments of Internal Medicine and Biomedical Sciences, Section of Geriatrics, University of Parma, Parma, Italy, **15** Sections of General Internal Medicine, Preventive Medicine, Cardiology, and Endocrinology, Diabetes, and Nutrition, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America, **16** The European Male Ageing Study, University of Manchester, Manchester, United Kingdom, **17** Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, United States of America, **18** Institute of Epidemiology II, Helmholtz Zentrum München, Neuherberg, Germany, **19** Department of Epidemiology, Erasmus MC Rotterdam, Rotterdam, The Netherlands, **20** Arthritis Research UK Epidemiology Unit, University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom, **21** Department of Physiology, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, **22** Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland, **23** Hebrew SeniorLife Institute for Aging Research and Harvard Medical School, Boston, Massachusetts, United States of America, **24** Clinical and Molecular Osteoporosis Research Unit, Department of Clinical Sciences and Department of Orthopaedics, Lund University, Skane University Hospital, Malmö, Sweden, **25** Department of Epidemiology and Prevention, Wake Forest University Health Sciences, Winston-Salem, North Carolina, United States of America, **26** Department of Medical Sciences, University of Uppsala, Uppsala, Sweden, **27** Institute of Human Genetics, Technische Universität München, München, Germany, **28** Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany, **29** Peninsula Medical School, University of Exeter, Exeter, United Kingdom, **30** University of Pittsburgh, Department of Epidemiology, Pittsburgh, Pennsylvania, United States of America, **31** Department of Psychiatry and EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands, **32** Medizinische Klinik Innenstadt, Ludwig-Maximilians-University, München, Germany, **33** Interfaculty Institute for Genetics and Functional Genomics, University of Greifswald, Greifswald, Germany, **34** Department of Medicine, University of Turku and Turku University Hospital, Turku, Finland, **35** Institute for Community Medicine, University of Greifswald, Greifswald, Germany, **36** Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians- University, München, Germany, **37** Klinikum Großhadern, München, Germany, **38** Institute of Epidemiology I, Helmholtz Zentrum München, Neuherberg, Germany, **39** Division of General Internal Medicine, Department of Medicine, University of California San Francisco, San Francisco, California, United States of America, **40** Department of Epidemiology and Biostatistics, Institute for Human Genetics, University of California San Francisco, San Francisco, California, United States of America, **41** Andrology Research Unit, Developmental and Regenerative Biomedicine Research Group, The University of Manchester, Manchester Academic Health Science Centre, Manchester Royal Infirmary, Manchester, United Kingdom, **42** Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Finland, **43** Department of Clinical Physiology, Turku University Hospital, Turku, Finland

Abstract

Testosterone concentrations in men are associated with cardiovascular morbidity, osteoporosis, and mortality and are affected by age, smoking, and obesity. Because of serum testosterone's high heritability, we performed a meta-analysis of genome-wide association data in 8,938 men from seven cohorts and followed up the genome-wide significant findings in one *in silico* (n = 871) and two *de novo* replication cohorts (n = 4,620) to identify genetic loci significantly associated with serum testosterone concentration in men. All these loci were also associated with low serum testosterone concentration defined as <300 ng/dl. Two single-nucleotide polymorphisms at the sex hormone-binding globulin *(SHBG)* locus (17p13-p12) were identified as independently associated with serum testosterone concentration (rs12150660, $p = 1.2 \times 10^{-41}$ and rs6258, $p = 2.3 \times 10^{-22}$). Subjects with ≥3 risk alleles of these variants had 6.5-fold higher risk of having low serum testosterone than subjects with no risk allele. The rs5934505 polymorphism near *FAM9B* on the X chromosome was also associated with testosterone concentrations ($p = 5.6 \times 10^{-16}$). The rs6258 polymorphism in exon 4 of *SHBG* affected SHBG's affinity for binding testosterone and the measured free testosterone fraction ($p < 0.01$). Genetic variants in the *SHBG* locus and on the X chromosome are associated with a substantial variation in testosterone concentrations and increased risk of low testosterone. rs6258 is the first reported *SHBG* polymorphism, which affects testosterone binding to SHBG and the free testosterone fraction and could therefore influence the calculation of free testosterone using law-of-mass-action equation.

Citation: Ohlsson C, Wallaschofski H, Lunetta KL, Stolk L, Perry JRB, et al. (2011) Genetic Determinants of Serum Testosterone Concentrations in Men. PLoS Genet 7(10): e1002313. doi:10.1371/journal.pgen.1002313

Editor: Gonçalo R. Abecasis, University of Michigan, United States of America

Received April 4, 2011; **Accepted** August 11, 2011; **Published** October 6, 2011

Funding: Framingham Heart Study (FHS): The FHS phenotype-genotype analyses for this work were supported by the National Institute of Aging (Genetics of Reproductive Life Period and Health Outcomes, R21AG032598; JM Murabito, KL Lunetta, D Karasik, DP Kiel, WV Zhuang). The Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine is supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study Contract No. N01-HC-25195 and its contract with Affymetrix for genotyping services (Contract No. N02-HL-6-4278). Sex hormone measurements were funded primarily by National Institute on Aging grant 1RO1AG31206 (PIs: S Bhasin and RS Vasan); additional support was provided by the Boston Claude D. Pepper Older Americans Independence Center (5P30AG031679) and a grant from the National Institute on Aging and the National Institute of Arthritis Musculoskeletal and Skin Diseases to DP Kiel (R01 AR/AG 41398). Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. A portion of this research was conducted using the Linux Cluster for Genetic Analysis (LinGA-II), funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Study of Health in Pomerania (SHIP): Computing resources have been made available by the Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities (HLRB project h1231). SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg – West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany, and the Federal State of Mecklenburg West Pomerania. The University of Greifswald is a member of the ''Center of Knowledge Interchange'' program of the Siemens AG. This work is also part of the research project Greifswald Approach to Individualized Medicine (GANI_MED). The GANI_MED consortium is funded by the Federal Ministry of Education and Research and the Ministry of Cultural Affairs of the Federal State of Mecklenburg – West Pomerania (03IS2061A). The testosterone reagents used were sponsored by Siemens Healthcare Diagnostics, Eschborn, formerly DPC Biermann GmbH, Bad Nauheim, Germany. Novo Nordisc provided partial grant support for the determination of serum samples and data analysis. Gothenburg Osteoporosis and Obesity Determinants (GOOD) Study: Financial support was received from the Swedish Research Council (K2010-54X-09894-19-3, 2006-3832, and K2010-52X-20229-05-3), the Swedish Foundation for Strategic Research, the ALF/LUA research grant in Gothenburg, the Lundberg Foundation, the Torsten and Ragnar Söderberg's Foundation, Petrus and Augusta Hedlunds Foundation, the Västra Götaland Foundation, the Göteborg Medical Society, the Novo Nordisk foundation, the Canadian Institutes of Health Research (MOP-15261), and the European Commission grant HEALTH-F2-2008-201865-GEFOS. We would like to acknowledge Maria Nethander at the genomics core facility at University of Gothenburg for statistical analyses. We would also like to thank Dr. Tobias A. Knoch, Luc V. de Zeeuw, Anis Abuseiris, and Rob de Graaf as well as their institutions the Erasmus Computing Grid, Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technology under grants #01 AK 803 A-H and # 01 IG 07015 G for access to their grid resources. Cooperative Research in the Region of Augsburg (KORA): The KORA research platform was initiated and financed by the Helmholtz Center Munich, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Part of this work was financed by the German National Genome Research Network (NGFN-2 and NGFNPlus: 01GS0823). Our research was supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ. This study was in part supported by a grant from the German Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DZD e.V.). Health, Aging, and Body Composition (Health ABC) Study: This study was supported by National Institute on Aging contracts N01-AG-6-2101, N01-AG-6-2103, and N01-AG-6-2106. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported (in part) by the Intramural Research Program of the NIH, National Institute on Aging. Rotterdam study (RS1): The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI) – Netherlands Consortium of Healthy Aging (NCHA) project nr. 050-060-810, and funding from the European Commision (HEALTH-F2-2008-201865, GEFOS; HEALTH-F2-2008-35627, TREAT-OA). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University Rotterdam; Netherlands Organisation for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); the Ministry of Education, Culture, and Science; the Ministry for Health, Welfare, and Sports; the European Commission (DG XII); and the Municipality of Rotterdam. We thank Pascal Arp, Mila Jhamai, Dr. Michael Moorhouse, Marijn Verkerk, and Sander Bervoets for their help in creating the GWAS database. The authors are grateful to the study participants, the staff from the Rotterdam Study, and the participating general practioners and pharmacists. We would like to thank Dr. Tobias A. Knoch, Luc V. de Zeeuw, Anis Abuseiris, and Rob de Graaf as well as their institutions the Erasmus Computing Grid, Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technology under grants #01 AK 803 A-H and # 01 IG 07015 G for access to their grid resources. Invecchiare in Chianti (InCHIANTI): The InCHIANTI study baseline (1998–2000) was supported as a ''targeted project'' (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336); the InCHIANTI Follow-up 1 (2001–2003) was funded by the U.S. National Institute on Aging (Contracts: N.1-AG-1-1 and N.1-AG-1-2111); the InCHIANTI Follow-ups 2 and 3 studies (2004–2010) were financed by the U.S. National Institute on Aging (Contract: N01-AG-5-0002); supported in part by the Intramural research program of the National Institute on Aging, National Institutes of Health, Baltimore, Maryland. European Male Ageing Study (EMAS): The EMAS is funded by the Commission of the European Communities Fifth Framework Programme ''Quality of Life and Management of Living Resources'' Grant QLK6-CT-2001-00258 and supported by funding from the UK Arthritis Research Campaign. The EMAS Principal Investigator is Professor Frederick Wu, MD; Dept of Endocrinology, Manchester Royal Infirmary, UK. The ''EMAS Study Group'' consists of the following people: Gyorgy Bartfai, Steven Boonen, Felipe Casanueva, Joseph D Finn, Gianni Forti, Aleksander Giwercman, Thang S Han, Kate L Holliday, Ilpo T Huhtaniemi, Krzysztof Kula, Michael EJ Lean, David M Lee, Terence W O'Neill, Neil Pendleton, Margus Punab, Stephen R Pye, Alan J Silman, Abdelouahid Tajar, Wendy Thomson, Dirk Vanderschueren, and Frederick CW Wu. The authors wish to thank the men who participated in the eight countries and the research/nursing staff in the eight centres: C Pott, Manchester, E Wouters, Leuven, M Nilsson, Malmö, M del Mar Fernandez, Santiago de Compostela,M Jedrzejowska, Lodz, H-M Tabo, Tartu, A Heredi, Szeged for their data collection and C Moseley, Manchester for data entry and project coordination. Dr. Vanderschueren is a senior clinical investigator supported by the Clinical Research Fund of the University Hospitals Leuven, Belgium. Dr. Boonen is a senior clinical investigator of the Fund for Scientific Research-Flanders, Belgium (F.W.O.–Vlaanderen). Dr. Boonen is holder of the Leuven University Chair in Metabolic Bone Diseases. The Osteoporotic Fractures in Men Study – Sweden (MrOS Sweden): Financial support was received from the Swedish Research Council (K2010-54X-09894-19-3, 2006-3832), the Swedish Foundation for Strategic Research, the ALF/LUA research grant in Gothenburg, the Lundberg Foundation, the Torsten and Ragnar Söderberg's Foundation, Petrus and Augusta Hedlunds Foundation, the Västra Götaland Foundation, the Göteborg Medical Society, the Novo Nordisk Foundation, and the European Commission grant HEALTH-F2-2008-201865-GEFOS. The Cardiovascular Risk in Young Finns Study (YFS): YFS has been financially supported by the Academy of Finland (grant no. 117797, 121584, and 126925), the Social Insurance Institution of Finland, Turku University Foundation, Tampere and Turku University Hospital Medical Funds, Emil Aaltonen Foundation (T.L), Juho Vainio Foundation, Paavo Nurmi Foundation, the Tampere Tubeculosis Foundation, the Orion-Farmos Research Foundation, Finnish Foundation of Cardiovascular Research, and Finnish Cultural Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: claes.ohlsson@medic.gu.se

◔ These authors contributed equally to this work.

¶ These authors were joint senior authors on this work.

Author Summary

Testosterone is the most important testicular androgen in men. Low serum testosterone concentrations are associated with cardiovascular morbidity, metabolic syndrome, type 2 diabetes mellitus, atherosclerosis, osteoporosis, sarcopenia, and increased mortality risk. Thus, there is growing evidence that serum testosterone is a valuable biomarker of men's overall health status. Studies in male twins indicate that there is a strong heritability of serum testosterone. Here we perform a large-scale genome-wide association study to examine the effects of common genetic variants on serum testosterone concentrations. By examining 14,429 men, we show that genetic variants in the sex hormone-binding globulin (*SHBG*) locus and on the X chromosome are associated with a substantial variation in serum testosterone concentrations and increased risk of low testosterone. The reported associations may now be used in order to better understand the functional background of recently identified disease associations related to low testosterone. Importantly, we identified the first known genetic variant, which affects SHBG's affinity for binding testosterone and the free testosterone fraction and could therefore influence the calculation of free testosterone. This finding suggests that individual-based SHBG-testosterone affinity constants are required depending on the genotype of this single-nucleotide polymorphism.

Introduction

Testosterone, the most important testicular androgen in men, is largely bound to two plasma proteins. Most of the circulating testosterone (~50–60%) is bound with high affinity to sex hormone-binding globulin (SHBG), while a smaller fraction (40–50%) is bound loosely to albumin, and 1–3% is unbound and termed free testosterone [1]. In prospective cohort studies, low serum testosterone concentrations are associated with cardiovascular morbidity, metabolic syndrome [2,3], dyslipidemia [4], hypertension [5], type 2 diabetes mellitus [6], stroke [7], atherosclerosis [8–10], osteoporosis, sarcopenia, and increased mortality risk [11–13]. Thus, there is growing evidence that serum testosterone is a valuable biomarker of men's overall health status. Since age, body mass index (BMI), and smoking are known to affect serum testosterone concentrations [14], we used these parameters as common set of covariates in all association models. Studies in male twins indicate that there is a strong heritability of serum testosterone, with genetic factors accounting for 65% of the variation in serum testosterone [15]. However, the genetic determinants of serum testosterone and the genetic risk factors for low concentrations are poorly understood. Given the current gap in knowledge of the genetic factors that contribute to the inter-individual variability in serum testosterone concentration in men we conducted a meta-analysis of genome-wide association studies (GWAS). This two-stage meta-analysis included data from 14,429 Caucasian men from 10 independent cohorts within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. In stage one, the discovery stage, genome-wide association data from seven cohorts were meta-analyzed (n = 8,938) and all genome-wide significant findings that fulfilled the criteria described in the methods section were followed up in the three replication cohorts: one *in silico* replication cohort (n = 871) and two replication cohorts with *de novo* genotyping (n = 4,620). All association analyses of the discovery stage were conducted both with and without additional adjustment for serum SHBG concentrations. Our primary aim was to identify genetic variants reproducibly associated with serum testosterone concentrations in men, evaluated as a continuous trait. We also assessed whether the lead single-nucleotide polymorphisms (SNPs) from the continuous trait analyses had a significant influence on the risk of having low serum testosterone, defined as <300 ng/dl [16]. This level is slightly lower than that suggested recently by Wu et al. [11 nmol/l = 317 ng/dl] as one of the clinical criteria for late onset hypogonadism [17].

Results

Meta-analyses of genome-wide association studies for autosomal SNPs

We performed a GWAS of serum testosterone concentrations, investigating ~2.5 million SNPs in 8,938 men of Caucasian ancestry, 18 to 98 years, from seven cohorts. Genome-wide significant SNPs were found in the discovery analysis at one locus on chromosome 17 (17p13-p12) using the criteria described in the methods. The strongest association was found for rs12150660 ($p = 1.9\times10^{-17}$), located 11.5 kb upstream of the major transcription start site of *sex hormone-binding globulin* (*SHBG*), with a minor allele frequency (MAF) of 23% (Table 1 [SNPs rs12150660 and rs6258], Figure 1A and Figures S1A, S2 and S3). Tests for independently associated SNPs with serum testosterone in this region revealed a second SNP, rs6258 ($p = 4.1\times10^{-14}$), which represents a missense (P→L) polymorphism located in exon 4 of *SHBG* (Table 1 [SNPs rs12150660 and rs6258], Figure 1B) and which had a MAF of 2%. Based on HapMap release 22 (CEU), the r^2 between rs12150660 and rs6258 was 0.004. To validate the independence of these two SNPs, conditional meta-analysis of the discovery cohorts including both rs12150660 and rs6258 in an additive genetic linear model adjusted for covariates was calculated. Because the associations remained significant and mostly unchanged (rs12150660, $p = 7.0\times10^{-14}$; rs6258, $p = 1.6\times10^{-13}$), both SNPs were independently associated with serum testosterone concentrations. No additional autosomal locus fulfilled the criteria for genome-wide significance.

Replication of autosomal hits

The associations of rs12150660 and rs6258 were confirmed in the three replication cohorts (*in silico* replication in YFS and *de novo* replication in MrOS Sweden and EMAS), demonstrating a combined p-value in the discovery and the replication cohorts of 1.2×10^{-41} and 2.3×10^{-22}, respectively (Table 1 [SNPs rs12150660 and rs6258]). Both SNPs showed considerable heterogeneity of results across the studies as measured by the I^2 statistic [18]. The I^2 values for the discovery meta-analysis using the untransformed total testosterone values were 76.7% and 81.6% for rs12150660 and rs6258, respectively. The heterogeneity was reduced to 39.3% and 75.5% for rs12150660 and rs6258, respectively, by meta-analysing the z-score based untransformed total testosterone values and to 30.9% and 78.0%, respectively, by meta-analysing the inverse-normal transformed testosterone values. For rs12150660, a substantial amount of heterogeneity could be explained by phenotypic variation among the cohorts, whereas for rs6258 one cohort (InCHIANTI) showed consistent opposite effect directions in all models used. To take into account this heterogeneity, we additionally calculated a random effects model for untransformed total testosterone values. The association for rs12150660 remained genome-wide significant in the combined discovery and replication stage meta-analysis, the association for

Table 1. Meta-analyses of discovery and replication cohorts.

SNPs rs12150660 and rs6258 (on chromosome 17 in *SHBG*) identified in GWAS for total testosterone														
	Discovery						**Replication**				**Combined**			
	A1/A2	**FREQ***	**beta**	**se**	**p**	**n**	**beta**	**se**	**p**	**n**	**beta**	**se**	**p**	**n**
Testosterone (ng/dl)														
rs12150660	T/G	0.23	26.4	3.1	1.9E-17	8938	38.8	3.6	2.3E-27	5429	31.8	2.3	1.2E-41	14367
rs6258	T/C	0.02	−74.7	9.9	4.1E-14	8938	−102.9	16.3	2.9E-10	5483	−82.3	8.5	2.3E-22	14421
SHBG (nmol/l)														
rs12150660	T/G	0.23	3.6	0.3	3.0E-42	8366	4.4	0.4	8.5E-36	5682	3.9	0.2	2.1E-75	14048
rs6258	T/C	0.02	−6.6	0.8	1.2E-15	8366	−9.5	1.3	6.7E-14	5733	−7.4	0.7	3.5E-27	14099
Testosterone (SHBG-adjusted)														
rs12150660	T/G	0.23	11.1	3.0	2.5E-04	8366	11.6	3.0	9.9E-05	5414	11.3	2.1	9.0E-08	13780
rs6258	T/C	0.02	−41.8	9.4	8.2E-06	8366	−33.2	13.8	1.6E-02	5467	−39.1	7.7	4.5E-07	13833
Calculated Free Testosterone (ng/dl)														
rs12150660	T/G	0.23	−0.1	0.1	9.6E-02	8366	0.1	0.1	1.6E-02	5414	0.0	0.0	3.9E-01	13780
rs6258	T/C	0.02	−0.2	0.2	3.2E-01	8366	−0.5	0.3	9.0E-02	5467	−0.3	0.2	6.5E-02	13833

SNP rs5934505 (on chromosome X near *FAM9B*) identified in GWAS for SHBG-adjusted total testosterone														
	Discovery						**Replication**				**Combined**			
	A1/A2	**FREQ***	**beta**	**se**	**p**	**n**	**beta**	**se**	**p**	**n**	**beta**	**se**	**p**	**n**
Testosterone (ng/dl)	C/T	0.26	14.1	3.2	1.1E-05	5067	27.2	6.0	5.4E-06	3816	17.0	2.8	1.7E-09	8883
SHBG (nmol/l)	C/T	0.26	−0.2	0.3	5.9E-01	4607	0.5	0.7	4.7E-01	4072	−0.1	0.3	8.5E-01	8679
Testosterone (SHBG-adjusted)	C/T	0.26	18.1	3.1	8.5E-09	4599	27.7	4.7	4.4E-09	3801	21.0	2.6	5.6E-16	8400
Calculated Free Testosterone (ng/dl)	C/T	0.26	0.4	0.1	4.0E-07	4607	0.6	0.1	8.7E-10	3801	0.5	0.1	6.7E-15	8408

Effects size is given per minor allele. All seven discovery cohorts (n = 8,938) were included in the GWAS of chromosomes 1–22 while only the two largest cohorts (FHS and SHIP. n = 5,067) had GWAS data available for the X chromosome. A1 = allele 1. A2 = allele 2. FREQ* = Frequency of allele 1. In the KORA cohort, testosterone was measured using plasma but the analyses after excluding KORA yielded similar results. Calculated free testosterone was calculated for all subjects with both testosterone and SHBG available by using a modified law of mass action equation. The concentrations of testosterone and SHBG and a fixed value for SHBG's dissociation constant were used in these calculations.

doi:10.1371/journal.pgen.1002313.t001

rs6258 reached genome-wide significance after excluding the InCHIANTI cohort (Table S3).

The genetic influence on low serum testosterone concentrations

In Table 2, the serum testosterone concentrations according to genotype are given for the three replication cohorts. As expected, mean serum testosterone concentrations were found to be lower in men with GG than in those with TT genotype for rs12150660. Similarly, men with the CT genotype for rs6258 had lower serum testosterone concentrations than those with CC genotype. The TT genotype of rs6258 was extremely rare and only found in two subjects in the replication cohorts. The two autosomal SNPs identified by GWAS had a significant influence on the risk of having low serum testosterone (serum testosterone <300 ng/dl) in both the discovery and the replication cohorts with a combined odds ratio (OR) per minor allele of 0.72 (95% CI, 0.65 – 0.79) and 2.7 (95% CI, 2.1 – 3.5) for rs12150660 and rs6258, respectively (Figure 2A). We analyzed the combined effect of the two SNPs on the risk of having low serum testosterone concentrations according to the number of combined risk alleles for rs12150660 (G) and rs6258 (T) in the three replication cohorts (MrOS Sweden, EMAS, and YFS). The risk of having low serum testosterone concentrations increased by the number of risk alleles with an OR of 1.62 (95% CI, 1.41 – 1.86) for each risk allele (Figure S4). Low serum testosterone concentrations were 6.5-times more prevalent in men with ≥3 risk alleles (30.1% prevalence of low serum testosterone) compared to men without any risk allele (4.6% prevalence of low serum testosterone; Figure 2B).

The role of SHBG in the observed associations

As SNP rs12150660 is located 11.5 kb upstream of *SHBG* and SNP rs6258 is non-synonymous and located in exon 4 of *SHBG*, we evaluated the influence of these polymorphisms on SHBG concentrations. Both of these polymorphisms demonstrated a significant association with SHBG concentrations in both the discovery and replication cohorts (Table 1 [SNPs rs12150660 and rs6258]). However, even after adjusting for SHBG concentrations, the associations between these two SNPs and serum testosterone concentrations were still significant ($p = 9.0 \times 10^{-8}$ for rs12150660 and $p = 4.5 \times 10^{-7}$ for rs6258). Free testosterone calculated using law-of-mass-action equation was not associated with either of the two polymorphisms (Table 1 [SNPs rs12150660 and rs6258]). As serum testosterone and SHBG are highly correlated (e.g., in MrOS Sweden $r_s = 0.53$), variations in SHBG concentration might have influenced the observed associations of serum testosterone with other non-SHBG-related loci. Therefore, we performed an additional SHBG-adjusted genome-wide meta-analysis among the discovery cohorts, wherein none of the non-SHBG-related autosomal SNPs reached genome-wide significance (Figure S1B).

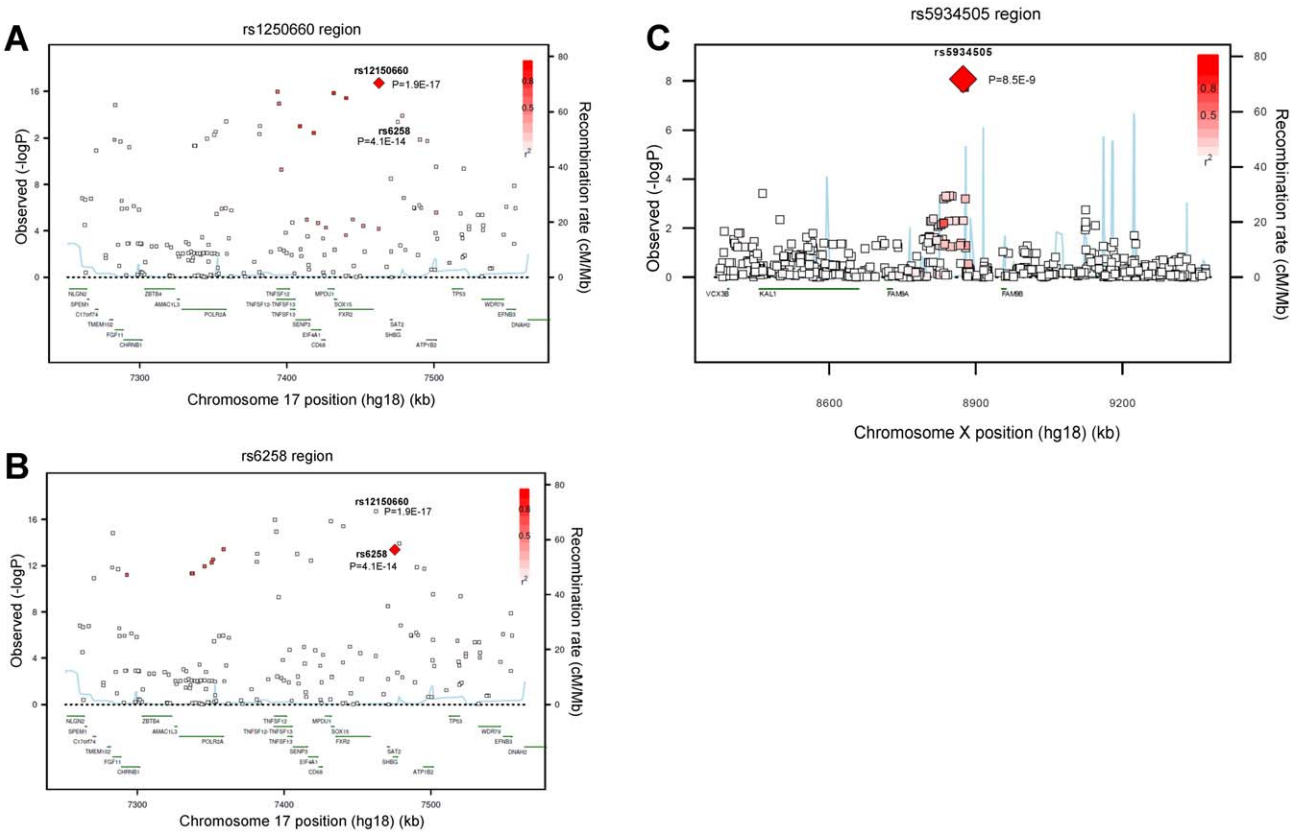


Figure 1. Regional association plots for single-nucleotide polymorphisms rs12150660, rs6258, and rs5934505. Regional association plot of the two independent signals on chromosome 17 with either (A) rs12150660 or (B) rs6258 indicated by red diamond to evaluate linkage with other single-nucleotide polymorphisms in the region. In addition, the association plot of the (C) rs5934505 signal on chromosome X is given. The r^2 is based on the CEU HapMap II samples. The blue line and right hand Y axis represent CEU HapMap II based recombination rates. (A) and (B) show the top SNPs of the inverse-variance weighted discovery stage meta-analysis of untransformed serum testosterone and (C) show the top SNP of the SHBG-adjusted serum testosterone using an imputation quality filter (observed/expected variance ratio) >0.4 at the individual cohort level during meta-analysis.
doi:10.1371/journal.pgen.1002313.g001

The rs6258 polymorphism affects SHBG binding affinity for testosterone and the measured free testosterone fraction

As rs6258 is non-synonymous (P156L) and located in exon 4 of *SHBG*, we evaluated the serum SHBG steroid-binding capacity of the different rs6258 genotypes. As shown in Figure S5, serum SHBG from CT but not CC subjects had a lower steroid-binding capacity than expected from values obtained by an SHBG immunoassay ($p = 0.003$). Therefore, we analyzed the SHBG affinity for testosterone using Scatchard plots of SHBG in serum of men with the rs6258 genotype (Figure 3A), and revealed (Figure 3B) a higher mean dissociation constant (Kd) indicative of a lower affinity in CT (Kd = 4.5 nM) and TT (Kd = 4.9 nM) individuals than in CC individuals (Kd = 2.8 nM). Recombinant SHBG corresponding to the T genotype demonstrated a higher dissociation constant (lower affinity) compared with recombinant SHBG corresponding to the C genotype (T genotype Kd 2.5 nM; C genotype Kd 1.2 nM, Figure 3C). In addition, the free testosterone fraction measured by an equilibrium dialysis method was 22% higher ($p = 1.4\times10^{-5}$) in serum from CT subjects than in serum from CC subjects (Figure 3D).

X chromosome analyses

Imputed values for X chromosome-located SNPs were available for the two larger discovery cohorts (SHIP and FHS; n = 5,067). We performed meta-analyses of imputed X chromosome SNPs for serum testosterone concentrations both with and without SHBG adjustment, revealing one genome-wide significant association for SNP rs5934505 ($p = 8.5\times10^{-9}$) in the SHBG-adjusted model (Table 1 [SNP rs5934505] and Figures S1B and S3). This SNP was confirmed in the two replication cohorts with *de novo* genotyping (MrOS Sweden $p = 3.6\times10^{-3}$; EMAS $p = 1.5\times10^{-7}$). Meta-analysis of discovery and replication cohorts resulted in a combined p-value of 5.6×10^{-16}. The rs5934505 SNP is located in a CNV-insertion area (Xp22), 145 kb upstream of the *family with sequence similarity 9, member A (FAM9A)* and 79 kb downstream of the *family with sequence similarity 9, member B* (*FAM9B*) (Figure 1C). In addition, rs5934505 is located 214 kb upstream of Kallmann syndrome 1 sequence (*KAL1*). SNP rs5934505 was associated with serum testosterone without SHBG-adjustment (combined p-value of 1.7×10^{-9}) and with free testosterone (combined p-value of 6.7×10^{-15}), but not with SHBG (Table 1 [SNP rs5934505]). The mean serum testosterone and calculated free testosterone but not SHBG concentrations were lower in men with T genotype than in those with C genotype for rs5934505 (Table 2).

Discussion

This GWAS revealed novel genetic variants that significantly affect circulating testosterone concentrations in men. The presence of three or more risk alleles for the two polymorphisms in the

Table 2. Serum sex steroids in the three replication cohorts according to rs12150660, rs6258, and rs5934505 genotype.

	SNPs identified in GWAS for total testosterone								SNP identified in GWAS for SHBG-adjusted testosterone		
	rs12150660				rs6258				rs5934505		
	GG	GT	TT	p-value	CC	CT	TT	p-value	C	T	p-value
EMAS	(n = 1310)	(n = 833)	(n = 152)		(n = 2261)	(n = 34)			(n = 410)	(n = 1120)	
Testosterone (ng/dl)	454±161	490±172	544±181	<0.001	474±169	358±104		<0.001	495±178	473±168	0.02
Calculated Free Testosterone (ng/dl)	8.47±2.53	8.53±2.53	8.84±2.85	0.15	8.52±2.56	8.14±2.14		0.39	9.00±2.65	8.45±2.49	<0.001
SHBG (nM)	39.6±17.1	45.2±20.4	51.6±20.8	<0.001	42.6±19.0	26.8±10.6		<0.001	42.4±20.5	42.8±18.9	0.69
MrOS Sweden	(n = 1317)	(n = 844)	(n = 123)		(n = 2245)	(n = 31)			(n = 530)	(n = 1765)	
Testosterone (ng/dl)	435±170	475±177	526±171	<0.001	456±174	331±125		<0.001	473±177	448±173	0.005
Calculated Free Testosterone (ng/dl)	7.98±3.07	8.30±3.16	8.75±2.99	0.005	8.16±3.08	7.59±2.72		0.31	8.54±3.27	8.03±3.03	0.001
SHBG (nM)	41.0±21.6	45.8±22.4	49.8±23.0	<0.001	43.5±22.0	24.3±12.3		<0.001	43.7±24.1	43.1±21.5	0.51
YFS	(n = 522)	(n = 329)	(n = 51)		(n = 852)	(n = 48)	(n = 2)				
Testosterone (ng/dl)	525±182	549±246	561±158	0.063	540±209	471±157	441±75	0.065	NA		
Calculated Free Testosterone (ng/dl)	11.89±5.30	12.30±8.92	11.57±2.46	0.71	12.04±6.90	11.80±3.42	11.55±1.23	0.80	NA		
SHBG (nM)	30.0±11.7	31.3±11.9	35.2±13.1	0.007	31.2±12.0	23.0±8.1	20.5±4.0	<0.001	NA		

NA = not available. Free testosterone was calculated for all subjects with both testosterone and SHBG available by using a modified law of mass action equation. The concentrations of testosterone and SHBG and a fixed value for SHBG's dissociation constant were used in these calculations.

doi:10.1371/journal.pgen.1002313.t002

SHBG loci resulted in markedly decreased testosterone concentrations compared to men with two or less risk alleles. Importantly, one of the identified genetic variations was associated with an alteration in SHBG's binding affinity for testosterone and the measured free testosterone fraction. In addition, we identified a locus on the X chromosome influencing serum testosterone concentrations. The genetic contribution of the polymorphisms to testosterone concentrations reported here is substantial; as a reference for comparison, the effect of these polymorphisms on testosterone concentrations in men is similar or greater than that for known risk factors such as age, smoking, and BMI [19,20].

These findings improve our understanding of the genetic factors that affect serum testosterone concentrations and contribute to the variation in testosterone concentrations in men. These polymorphisms may assist in the identification of men at risk of low serum testosterone, although the clinical usefulness of these findings remains to be established. As rs12150660 and rs6258 were strongly associated with SHBG concentrations, both SNPs may at least partly affect total testosterone concentrations by modulating SHBG concentrations. Our findings that rs6258 substantially affects SHBG binding affinity and the measured free testosterone fraction raise questions about the use of a single consensus value for SHBG's dissociation constant in the law of mass action equations used to calculate free testosterone concentrations. As emphasized by the Endocrine Society's expert panel on androgen deficiency syndromes, low testosterone concentrations alone should not necessarily be viewed as evidence of androgen deficiency [16]. Whether rs593405 near the *FAM9B* and *KAL1* genes on Xp22 renders men susceptible to the increased risk of androgen deficiency remains to be determined. Further studies are required to determine the impact of these genetic variations on sex steroid-related disorders, including osteoporosis, cardiovascular diseases, prostate cancer, and male infertility [21].

Our studies add to the evidence that genetic variations within the *SHBG* gene may explain some of the inter-individual differences in SHBG concentrations. Our finding that SNP rs6258 results in the production of an *SHBG* variant with reduced affinity for testosterone provides an explanation for the association between rs6258 and low serum testosterone concentrations. This is the first described genetic variant associated with altered SHBG binding for testosterone. As rs6258 is non-synonymous (P156L), located in exon 4 of *SHBG* and associated with altered SHBG binding for testosterone and free testosterone fraction, rs6258 is likely a functional polymorphism with impact on testosterone binding to SHBG as well as testosterone bioavailability and action at target tissue level.

The SNP rs12150660 that is strongly associated with testosterone concentrations is located 11.5 kb upstream of the coding sequence for SHBG mRNA production in the liver. However, it still resides within the human *SHBG* locus because several other alternative exon 1 sequences are located up to ~13 kb upstream of the exon 1 sequence that encodes the secretion signal polypeptide of the SHBG precursor in the liver [22]. There are no obvious nuclear protein binding sites within the sequences spanning SNP rs12150660, and it remains to be determined whether this SNP disrupts a cis-element that directly influences *SHBG* transcription. We have found that rs12150660 is in strong LD ($r^2 = 0.89$) with another common SNP (rs1799941) in the *SHBG* proximal promoter that was shown to be associated with serum SHBG concentrations [23–25]. Thus, it is highly likely that only one of these polymorphisms is actually functional and therefore both SNPs represent the same signal. It should also be noted that rs1799941 is linked to the number of TAAAA repeats within an Alu sequence upstream of *SHBG* promoter [26] and that the rs1799941 (A allele) is linked with the presence of six TAAAA repeats in this location which has been reported to be associated with higher SHBG concentrations [27]. In addition, while there does not appear to be any putative transcriptional factor binding sites with the sequence comprising rs12150660, it remains to be determined whether rs12150660 or these other associated SNPs in

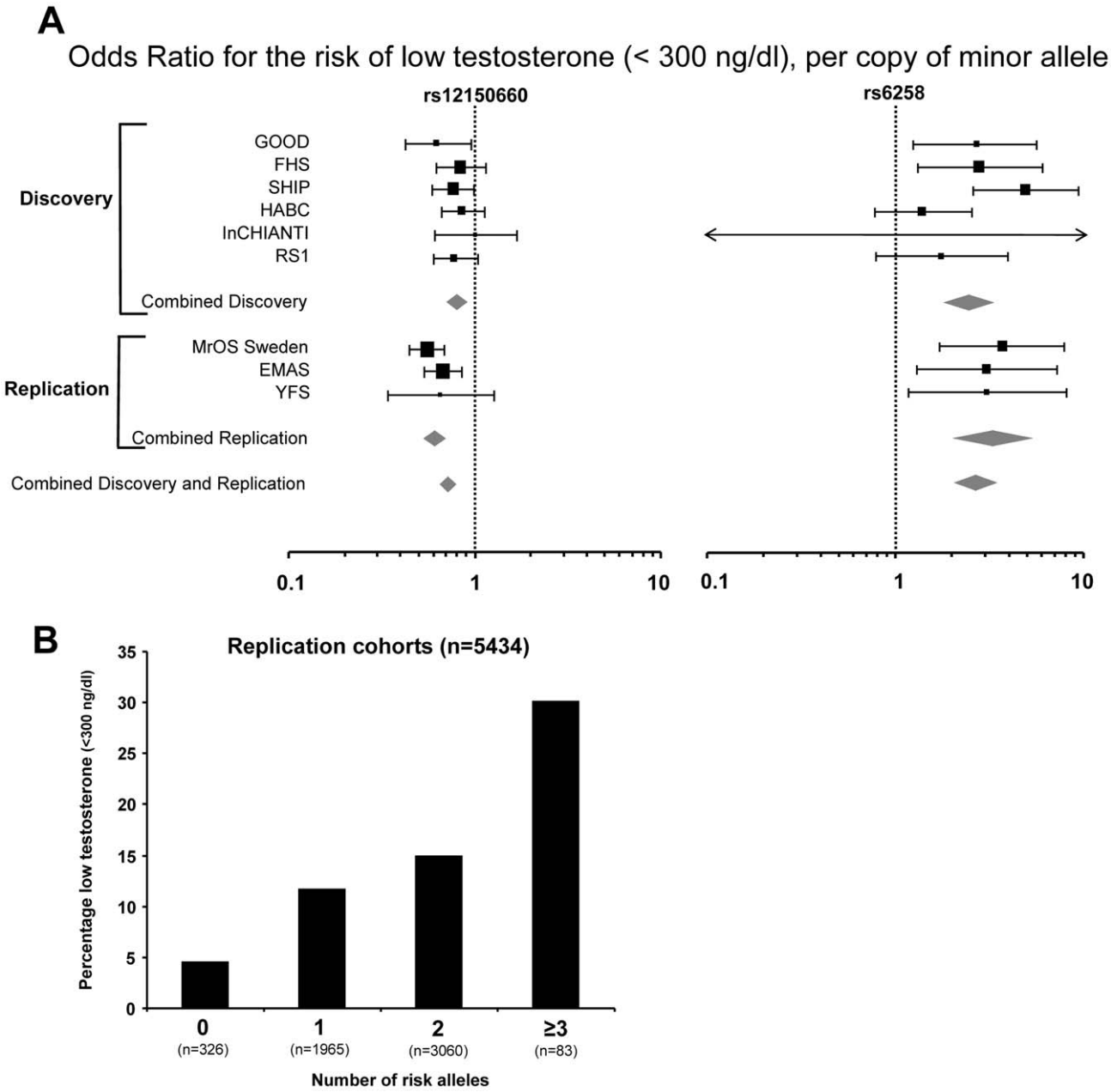


Figure 2. The genetic influence on low serum testosterone concentrations. (A) Odds ratio (OR) for risk of low serum testosterone concentrations (serum testosterone <300 ng/dl), per copy of minor allele. Summary estimates of the OR and their 95% confidence intervals (CI) are given. The size of the data markers is proportional to the weight (inverse of the variance) of each study. Combined discovery (n = 8,030, low serum testosterone 13%; KORA was not included as testosterone was analyzed in plasma rather than in serum, combined replication (n = 5,504, low serum testosterone 13%), and combined discovery and replication (n = 13,534, low serum testosterone 13%). (B) Percentage of men with low serum testosterone concentrations (serum testosterone <300 ng/dl), according to the number of combined risk alleles for rs12150660 (G) and rs6258 (T) in the three replication cohorts (MrOS Sweden, EMAS, and YFS). Only two individuals in the three replication cohorts had four risk alleles and therefore individuals with three and four risk alleles were pooled into one group with ≥3 risk alleles.
doi:10.1371/journal.pgen.1002313.g002

the *SHBG* gene are functionally important or simply represent proxies of SHBG and testosterone concentrations in men.

Our meta-analyses of imputed X chromosome SNPs revealed one genome-wide significant association for SNP rs5934505, located in a CNV-insertion area (Xp22), 145 kb upstream of *family with sequence similarity 9, member A* (*FAM9A*) and 79 kb downstream of *family with sequence similarity 9, member B* (*FAM9B*). Both genes, *FAM9A* and *FAM9B*, are expressed exclusively in the testis [28] and described here for the first time to be associated with total as well as free testosterone concentrations. rs5934505 is located 214 kb upstream of Kallmann syndrome 1 sequence (*KAL1*). Although the Kallmann syndrome, a type of hypogonadotropic hypogonadism associated with anosmia and other congenital anomalies, has been linked to mutations in the *KAL1* gene on the X chromosome, only 11–14% of Caucasian patients with hypogonadotropic hypogonadism have detectable *KAL1* mutations [29], reflecting the considerable genetic heterogeneity of this syndrome.

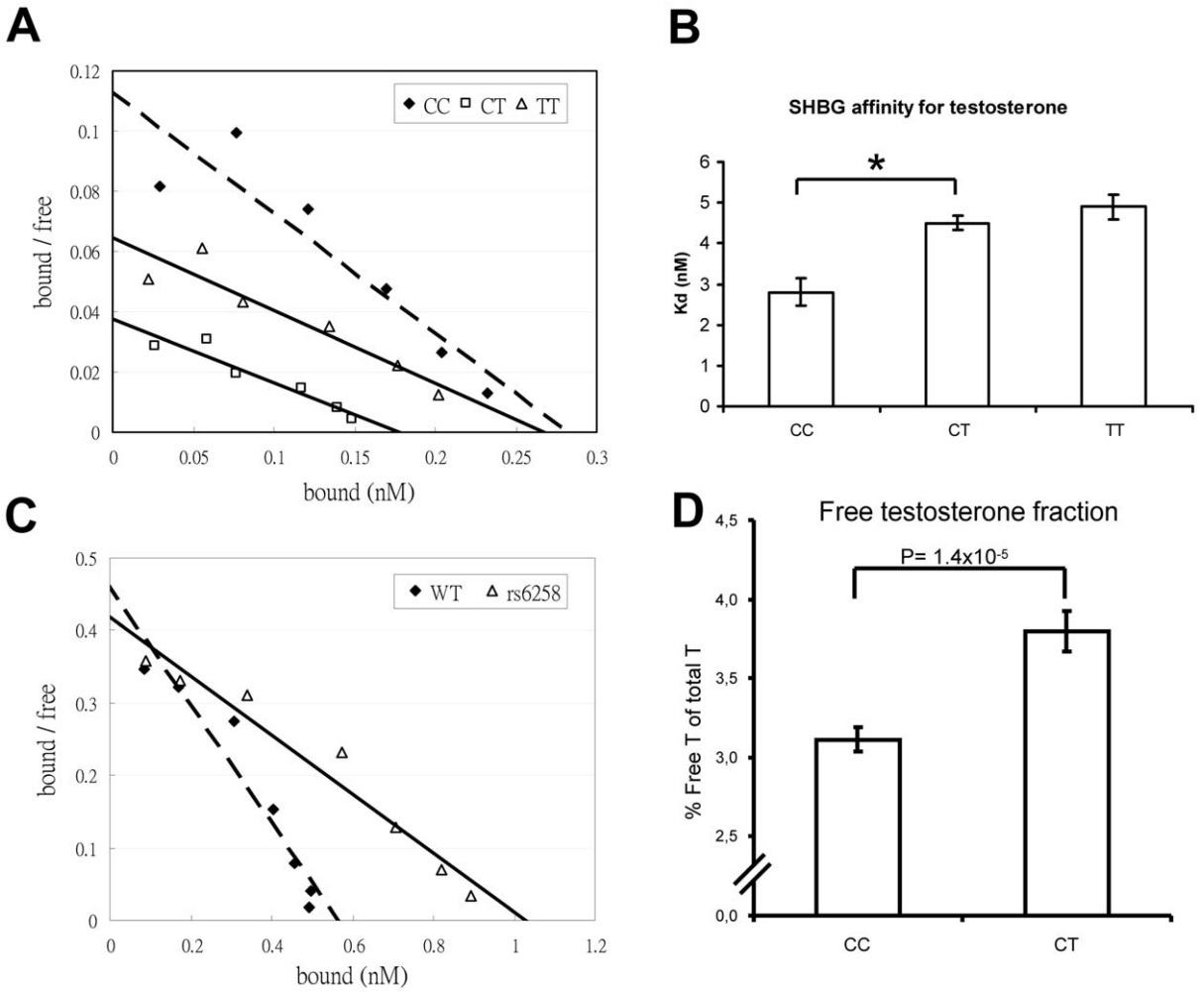


Figure 3. SHBG affinity for testosterone. (A and B) Scatchard plots of SHBG binding affinity for testosterone in serum samples according to rs6258 genotype. (A) Representative Scatchard plots of serum SHBG binding to [^{3}H]testosterone. Serum from individuals homozygous for the wild-type SHBG allele (CC dashed line) or the rs6258 SNP (TT, solid line), or heterozygous for these alleles (CT, solid line). (B) Dissociation constant (Kd) of serum SHBG according to rs6258 genotype (CC, n = 4 subjects; CT, n = 4 subjects; TT [rare variant] n = 1 and the variation for the TT subject is derived from three separate analyses). (*) p = 0.001. Values are means ± SEM. (C) Representative Scatchard plots of recombinant SHBG binding to [^{3}H]testosterone. Recombinant wild type (= WT, C genotype; dashed line) or rs6258 (T genotype; solid line) SHBG expressed by CHO cells was diluted 1:10 and subjected to Scatchard analysis, as in panel A. (D) Free testosterone fraction in serum measured by an equilibrium dialysis method according to rs6258 genotype (CC, n = 87 subjects; CT, n = 32 subjects). Values are means ± SEM.
doi:10.1371/journal.pgen.1002313.g003

The strengths of our study include a discovery sample size of 8,938 men, which allowed us at the threshold $\alpha = 5 \times 10^{-8}$,a 90% power to detect SNPs accounting for 0.5% of the total variance in serum testosterone concentrations, and 99% power to detect SNPs accounting for 1% of the total variance. The SNPs rs12150660, rs6258, and rs5934505 explained 2.3%, 0.9%, and 0.6%, respectively, of the variance in serum testosterone concentrations when evaluated in the MrOS Sweden replication cohort. Future meta-analyses including larger samples will probably reveal additional loci associated with serum testosterone. Further research into the functional significance of these variants will be needed to enable the translation of these findings into the mechanisms of sex steroid-related diseases and strategies for risk assessment. As the causal or etiological role of these polymorphisms in the genesis of low testosterone has not been established, the reported polymorphisms associated with low serum testosterone concentration may be viewed currently as risk markers rather than causal risk factors.

In conclusion, genetic variants in the *SHBG* locus and on the X chromosome are associated with a substantial variation in testosterone concentrations and increased risk of low testosterone in men. Further studies are needed to determine the impact of these genetic variations on sex hormone-related disorders. rs6258 is the first reported *SHBG* polymorphism, which affects testosterone binding to SHBG and the free testosterone fraction and could therefore influence the calculation of free testosterone using law-of-mass-action equation.

Methods

Study samples and genotyping

The discovery stage of the GWAS included 8,938 Caucasian men of European descent drawn from seven epidemiological cohorts: the Framingham Heart Study (FHS), the Study of Health in Pomerania (SHIP), the Gothenburg Osteoporosis and Obesity Determinants (GOOD) study, the Cooperative Health Research in

the Region of Augsburg (KORA) study, the Health, Aging and Body Composition (HEALTH ABC) study, the Rotterdam Study (RS1), and the Invecchiare in Chianti (InCHIANTI) (Table S1). The replication stage consisted of 4,620 men from two epidemiological cohorts (the European Male Ageing Study [EMAS] and the Osteoporotic Fractures in Men [MrOS] Sweden study) for *de novo* genotyping of the top SNPs and one additional cohort (the Young Finns Study, [YFS, n = 871]) with genome-wide association data available and joining the study after stage one was finished for *in silico* replication (Table S2).

Exclusion criteria included chemical or surgical castration and/or medications affecting sex hormones such as steroid 5-alpha reductase inhibitors, and sex hormone antagonists. All studies were approved by local ethics committees and all participants provided written informed consent. Characteristics of the study samples and detailed descriptions of the participating cohorts, genotyping methods, quality control, and imputation procedures are provided in Text S1.

Genotyping and statistical analyses

Altogether, ~2.5 million SNPs, imputed using the HapMapII CEU population, were tested for association with serum testosterone in the discovery stage. Genome-wide association analyses using an additive genetic linear regression model adjusted for age, BMI, and current smoking were conducted twice within each of the discovery cohorts using serum testosterone expressed as ng/dl, as well as inverse-normal transformed serum testosterone as outcomes.

To examine the robustness of the discovery results and to reduce the risk of spurious associations due to possible testosterone measurement heterogeneity between the individual cohorts, three different types of meta-analyses were performed in the discovery stage: 1) an inverse-variance weighted fixed effect model; 2) a z-score based analysis of the untransformed serum testosterone concentrations; and 3) a z-score based meta-analysis of the inverse-normal transformed values. Model 1) was used as main analysis since it allowed the computation of effect estimates, whereas the other two analysis models were used for verification and quality control checks of the main findings. All meta-analyses were performed using METAL (www.sph.umich.edu/csg/abecasis/metal/). The random effects model of the two *SHBG* locus SNPs was calculated using the R-package *metafor* (www.r-project.org). Imputed genotypes were analyzed in all cohorts taking the genotype uncertainties into account. Genomic control was applied to each individual cohort's results and to the discovery stage meta-analysis to correct p-values for potential effects of mild population stratification. The estimated genomic control lambda was low for both the individual cohorts (range of λ_{GC}: 1.00–1.07) and the meta-analyses (range of λ_{GC}: 1.01–1.02), suggesting little residual confounding due to population stratification (Figure S2).

To reduce the variance on serum testosterone induced by SHBG concentration, the GWAS included a genome-wide test for association of untransformed serum testosterone concentrations adjusted for age, BMI, current smoking, SHBG and $SHBG^2$ concentrations, again using both an inverse-variance weighted fixed effect as main analysis and a z-score based meta-analysis for quality control purposes.

A threshold of $p<5\times10^{-8}$ was established *a priori* as the level for genome-wide significance in the discovery analyses [30]. SNPs that reached genome-wide significance in the inverse-variance weighted meta-analysis of untransformed serum testosterone concentrations with or without adjustment for SHBG and which had association results in at least five of the seven cohorts (for chr X: two cohorts with data available) were selected for further analyses. Notably, all autosomal SNPs that fulfilled these criteria also reached genome-wide significance in the other two types of meta-analyses. From these SNPs, all independent SNPs were taken to the replication stage.

We also assessed whether the lead SNPs from the continuous trait analyses were associated with low serum testosterone concentration (defined as <300 ng/dl [16]; this level is slightly lower than that suggested recently by Wu et al [11 nmol/l = 317 ng/dl] [17]) by binary logistic regression including the same covariates in the model used for the main analysis and meta-analyzing the within-cohort results using inverse-variance weighted fixed-effect model. The KORA cohort was not included in the meta-analyses of low serum testosterone as testosterone was measured using plasma in this cohort.

We determined the number of low serum testosterone concentration risk alleles (0 to 4) for the two lead SNPs of the *SHBG* locus in each individual and assessed the risk of low serum testosterone concentrations in the three replication cohorts (MrOS Sweden, EMAS, and YFS) using a trend test. Since only two subjects in the replication cohorts had four risk alleles, individuals having three and four risk alleles were grouped into one category to obtain more reliable effect estimates during the subsequent analyses. Details of test for independence, SHBG related analysis of the top SNPs and quality control steps performed can be found in Text S1.

Sex hormone measurements

Methods for the measurement of serum testosterone and SHBG are given in Text S1. Calculated free testosterone was for all subjects with both testosterone and SHBG available (n = 13833; Table 1 and Table 2) calculated by using a modified law of mass action equation, as described by Mazer [31]. The concentrations of testosterone and SHBG and a fixed value for SHBG's dissociation constant were used in these calculations.

Free testosterone fraction

Free testosterone fraction was measured by an equilibrium dialysis method in 87 subjects with the CC genotype and 32 subjects with the CT genotype of rs6258 (Figure 3D) [32]. Detailed description of the free testosterone fraction measurements is provided in Text S1.

Sex hormone-binding globulin assays

In experiments evaluating SHBG binding capacity, serum SHBG concentrations were determined by two-site immunofluorometric assay (PerkinElmer Life Sciences, Turku, Finland) [33], or by a steroid-binding capacity assay [34]. For steroid-binding assays, serum samples were pre-incubated with dextran-coated charcoal (DCC) to remove endogenous steroids, prior to incubation with either [^{3}H]5α-dihydrotestosterone ([^{3}H]DHT; specific activity 50 Ci/mmol) or [^{3}H]testosterone (specific activity 40 Ci/mmol), bound from free [^{3}H]steroid were separated using DCC as the separation reagent [34]. The steroid-binding properties of SHBG in diluted serum samples or tissue culture medium were determined by Scatchard analysis [34]. For the expression of SHBG protein, wild type (corresponding to the C genotype of rs6258) and rs6258 (corresponding to the T genotype of rs6258) SHBG cDNAs in the pRC/CMV expression vector were transfected into CHO cells, and G418 was used for selection of stably transfected cells. At near confluence, cells were washed with PBS and cultured in serum-free SFM4CHO medium (Thermo Scientific HyClone, Logan, UT) for four days before the SHBG-containing medium was harvested.

Supporting Information

Figure S1 Manhattan plots giving genome-wide $-\log_{10}$ p-value according to chromosomal location for inverse-variance weighted meta-analysis of untransformed serum testosterone (A) and SHBG-adjusted serum testosterone (B) using an imputation quality filter (observed/expected variance ratio) >0.4 at the individual cohort level during meta-analysis. All seven discovery cohorts (n = 8,938) were included in the GWAS of chromosomes 1–22 while only the two largest cohorts (FHS and SHIP, n = 5,067) had GWAS data available for the X chromosome.
(PDF)

Figure S2 Quantile-quantile plot of the genome-wide association results of the inverse-variance weighted meta-analysis of untransformed serum testosterone including all SNPs (black) and after removal of the SNPs of the *SHBG* locus (blue).
(PDF)

Figure S3 Associations for (A) rs12150660 and (B) rs6258 with testosterone and for (C) rs5934505 with SHBG-adjusted testosterone. Effects sizes are given per minor allele. Beta estimates and their 95% confidence intervals are given. The size of the data markers is proportional to the weight (inverse of the variance) of each study.
(PDF)

Figure S4 Risk of low serum testosterone concentrations (serum testosterone <300 ng/dl), according to the number of combined risk alleles for rs12150660 (G = risk allele) and rs6258 (T = risk allele) in the three replication cohorts (MrOS Sweden, EMAS, and YFS). Bars indicate 95% confidence intervals. Only two individuals in the three replication cohorts had four risk alleles and therefore individuals with three and four risk alleles were pooled into one group with ≥3 risk alleles. Two risk allele counts were used as reference, since this is the most prevalent amount among the cohorts.
(PDF)

Figure S5 Subjects heterozygous for the *SHBG* allele containing an rs6258 SNP have lower serum SHBG steroid-binding capacity (Y-axis) when compared to the concentrations of SHBG measured by immunoassay (X-axis). Serum SHBG concentrations from 10 individuals homozygous for the wild type *SHBG* allele (CC, dashed regression line $r^2 = 0.872$) or heterozygous for the rs6258 variant *SHBG* allele (CT, solid regression line $r^2 = 0.866$) were measured by a time-resolved immunofluorometric assay[33], and a steroid-binding capacity assay using [^{3}H]DHT as the labelled ligand.[34]
(PDF)

Table S1 Characteristics of 14,429 men from 10 cohorts included in the genome-wide association study meta-analysis.
(PDF)

Table S2 Additional genotyping information for the 10 cohorts included in the genome-wide association study meta-analysis.
(PDF)

Table S3 Meta Analysis of untransformed total testosterone using Random Effect Model.
(PDF)

Text S1 Supplemental methods.
(DOC)

Acknowledgments

We are indebted to the participants and staff of all the studies for their important contributions. Full individual study Acknowledgments are listed in Text S1.

Author Contributions

Conceived and designed the experiments: C Ohlsson, KL Lunetta, JRB Perry, T Lehtimäki, M Maggio, L Ferrucci, A Hofman, M Kähönen, D Karasik, DP Kiel, T Meitinger, D Melzer, RS Vasan, M Reincke, A Teumer, AG Uitterlinden, J Viikari, HE Wichmann, O Raitakari, M Bidlingmaier, TB Harris, JM Murabito, S Bhasin, L Vandenput, R Haring. Performed the experiments: C Ohlsson, H Wallaschofski, T Lehtimäki, M Maggio, L Ferrucci, M Heier, A Hofman, KL Holliday, J-O Jansson, M Kähönen, MK Karlsson, Ö Ljunggren, M Lorentzon, L-P Lyytikäinen, D Mellström, D Melzer, M Nauck, M Nilsson, B Penninx, RS Vasan, M Reincke, F Rivadeneira, AG Uitterlinden, J Ulloor, J Viikari, H Völzke, HE Wichmann, T-S Wu, O Raitakari, A Eriksson, M Bidlingmaier, FH de Jong, JM Murabito, S Bhasin. Analyzed the data: KL Lunetta, L Stolk, JRB Perry, A-K Petersen, KL Holliday, Y Liu, SR Pye, A Tajar, A Teumer, U Völker, WV Zhuang, M Bidlingmaier. Wrote the paper: C Ohlsson, H Wallaschofski, KL Lunetta, L Stolk, JRB Perry, A Koster, A-K Petersen, J Eriksson, T Lehtimäki, IT Huhtaniemi, GL Hammond, M Maggio, AD Coviello, L Ferrucci, M Heier, A Hofman, KL Holliday, J-O Jansson, M Kähönen, D Karasik, DP Kiel, Y Liu, L-P Lyytikäinen, I Miljkovic, M Nilsson, B Penninx, SR Pye, RS Vasan, M Reincke, F Rivadeneira, A Tajar, A Teumer, AG Uitterlinden, J Viikari, E Ziv, FCW Wu, O Raitakari, A Eriksson, M Bidlingmaier, A Murray, FH de Jong, JM Murabito, S Bhasin, L Vandenput, R Haring.

References

1. Kaufman JM, Vermeulen A (2005) The decline of androgen levels in elderly men and its clinical and therapeutic implications. Endocr Rev 26: 833–876.
2. Haring R, Volzke H, Felix SB, Schipf S, Dorr M, et al. (2009) Prediction of metabolic syndrome by low serum testosterone levels in men: results from the study of health in Pomerania. Diabetes 58: 2027–2031.
3. Kupelian V, Page ST, Araujo AB, Travison TG, Bremner WJ, et al. (2006) Low sex hormone-binding globulin, total testosterone, and symptomatic androgen deficiency are associated with development of the metabolic syndrome in nonobese men. J Clin Endocrinol Metab 91: 843–850.
4. Haring R, Baumeister SE, Völzke H, Dorr M, Felix SB, et al. (2011) Prospective Association of Low Total Testosterone Concentrations with an Adverse Lipid Profile and Increased Incident Dyslipidemia. Eur J Cardiovasc Prev Rehabil 18: 86–96.
5. Torkler S, Wallaschofski H, Baumeister SE, Völzke H, Dörr M, et al. (2011) Inverse Association Between Total Testosterone Concentrations, Incident Hypertension, and Blood Pressure. Aging Male 14: 176–182.
6. Vikan T, Schirmer H, Njolstad I, Svartberg J (2010) Low testosterone and sex hormone-binding globulin levels and high estradiol levels are independent predictors of type 2 diabetes in men. Eur J Endocrinol 162: 747–754.
7. Yeap BB, Hyde Z, Almeida OP, Norman PE, Chubb SA, et al. (2009) Lower testosterone levels predict incident stroke and transient ischemic attack in older men. J Clin Endocrinol Metab 94: 2353–2359.
8. Svartberg J, von Muhlen D, Mathiesen E, Joakimsen O, Bonaa KH, et al. (2006) Low testosterone levels are associated with carotid atherosclerosis in men. J Intern Med 259: 576–582.
9. Vikan T, Johnsen SH, Schirmer H, Njolstad I, Svartberg J (2009) Endogenous testosterone and the prospective association with carotid atherosclerosis in men: the Tromso study. Eur J Epidemiol 24: 289–295.
10. Tivesten A, Mellstrom D, Jutberger H, Fagerberg B, Lernfelt B, et al. (2007) Low serum testosterone and high serum estradiol associate with lower extremity peripheral arterial disease in elderly men. The MrOS Study in Sweden. J Am Coll Cardiol 50: 1070–1076.
11. Haring R, Volzke H, Steveling A, Krebs A, Felix SB, et al. (2010) Low serum testosterone levels are associated with increased risk of mortality in a population-based cohort of men aged 20-79. Eur Heart J 31: 1494–1501.
12. Laughlin GA, Barrett-Connor E, Bergstrom J (2008) Low serum testosterone and mortality in older men. J Clin Endocrinol Metab 93: 68–75.
13. Tivesten A, Vandenput L, Labrie F, Karlsson MK, Ljunggren O, et al. (2009) Low serum testosterone and estradiol predict mortality in elderly men. J Clin Endocrinol Metab 94: 2482–2488.
14. Wu FC, Tajar A, Pye SR, Silman AJ, Finn JD, et al. (2008) Hypothalamic-pituitary-testicular axis disruptions in older men are differentially linked to age and modifiable risk factors: the European Male Aging Study. J Clin Endocrinol Metab 93: 2737–2745.
15. Bogaert V, Taes Y, Konings P, Van Steen K, De Bacquer D, et al. (2008) Heritability of blood concentrations of sex-steroids in relation to body composition in young adult male siblings. Clin Endocrinol (Oxf) 69: 129–135.
16. Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, et al. (2010) Testosterone therapy in men with androgen deficiency syndromes: an

Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 95: 2536–2559.
17. Wu FC, Tajar A, Beynon JM, Pye SR, Silman AJ, et al. (2010) Identification of late-onset hypogonadism in middle-aged and elderly men. N Engl J Med 363: 123–135.
18. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. Bmj 327: 557–560.
19. Haring R, Ittermann T, Volzke H, Krebs A, Zygmunt M, et al. (2010) Prevalence, incidence and risk factors of testosterone deficiency in a population-based cohort of men: results from the study of health in Pomerania. Aging Male 13: 247–257.
20. Travison TG, Araujo AB, Kupelian V, O'Donnell AB, McKinlay JB (2007) The relative contributions of aging, health, and lifestyle factors to serum testosterone decline in men. J Clin Endocrinol Metab 92: 549–555.
21. Basaria S, Coviello AD, Travison TG, Storer TW, Farwell WR, et al. (2010) Adverse events associated with testosterone administration. N Engl J Med 363: 109–122.
22. Pinos T, Barbosa-Desongles A, Hurtado A, Santamaria-Martinez A, de Torres I, et al. (2009) Identification, characterization and expression of novel Sex Hormone Binding Globulin alternative first exons in the human prostate. BMC Mol Biol 10: 59.
23. Ahn J, Schumacher FR, Berndt SI, Pfeiffer R, Albanes D, et al. (2009) Quantitative trait loci predicting circulating sex steroid hormones in men from the NCI-Breast and Prostate Cancer Cohort Consortium (BPC3). Hum Mol Genet 18: 3749–3757.
24. Eriksson AL, Lorentzon M, Mellstrom D, Vandenput L, Swanson C, et al. (2006) SHBG gene promoter polymorphisms in men are associated with serum sex hormone-binding globulin, androgen and androgen metabolite levels, and hip bone mineral density. J Clin Endocrinol Metab 91: 5029–5037.
25. Perry JR, Weedon MN, Langenberg C, Jackson AU, Lyssenko V, et al. Genetic evidence that raised sex hormone binding globulin (SHBG) levels reduce the risk of type 2 diabetes. Hum Mol Genet 19: 535–544.
26. Hogeveen KN, Talikka M, Hammond GL (2001) Human sex hormone-binding globulin promoter activity is influenced by a (TAAAA)n repeat element within an Alu sequence. J Biol Chem 276: 36383–36390.
27. Xita N, Tsatsoulis A, Chatzikyriakidou A, Georgiou I (2003) Association of the (TAAAA)n repeat polymorphism in the sex hormone-binding globulin (SHBG) gene with polycystic ovary syndrome and relation to SHBG serum levels. J Clin Endocrinol Metab 88: 5976–5980.
28. Martinez-Garay I, Jablonka S, Sutajova M, Steuernagel P, Gal A, et al. (2002) A new gene family (FAM9) of low-copy repeats in Xp22.3 expressed exclusively in testis: implications for recombinations in this region. Genomics 80: 259–267.
29. Oliveira LM, Seminara SB, Beranova M, Hayes FJ, Valkenburgh SB, et al. (2001) The importance of autosomal genes in Kallmann syndrome: genotype-phenotype correlations and neuroendocrine characteristics. J Clin Endocrinol Metab 86: 1532–1538.
30. Pe'er I, Yelensky R, Altshuler D, Daly MJ (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 32: 381–385.
31. Mazer NA (2009) A novel spreadsheet method for calculating the free serum concentrations of testosterone, dihydrotestosterone, estradiol, estrone and cortisol: with illustrative examples from male and female populations. Steroids 74: 512–519.
32. Sinha-Hikim I, Arver S, Beall G, Shen R, Guerrero M, et al. (1998) The use of a sensitive equilibrium dialysis method for the measurement of free testosterone levels in healthy, cycling women and in human immunodeficiency virus-infected women. J Clin Endocrinol Metab 83: 1312–1318.
33. Niemi S, Maentausta O, Bolton NJ, Hammond GL (1988) Time-resolved immunofluorometric assay of sex-hormone binding globulin. Clin Chem 34: 63–66.
34. Hammond GL, Lahteenmaki PL (1983) A versatile method for the determination of serum cortisol binding globulin and sex hormone binding globulin binding capacities. Clin Chim Acta 132: 101–110.

ELSEVIER

Journal of Substance Abuse Treatment 41 (2011) 233–242

Journal of Substance Abuse Treatment

Regular article

Substance abuse treatment utilization among adults living with HIV/AIDS and alcohol or drug problems

John Orwat, (Ph.D.)[a,*], Richard Saitz, (M.D., M.P.H.)[b,f], Christopher P. Tompkins, (Ph.D.)[d], Debbie M. Cheng, (Sc.D.)[b,e], Michael P. Dentato, (M.S.W.)[a], Jeffrey H. Samet, (M.D., M.A., M.P.H.)[b,c]

[a]*Loyola University Chicago, Chicago, IL, USA*
[b]*Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA*
[c]*Department of Community Health Sciences, Boston University School of Public Health, Boston, MA, USA*
[d]*Brandeis University, Heller School for Social Policy & Management, Waltham, MA, USA*
[e]*Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA*
[f]*Department of Epidemiology and Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA, USA*

Received 10 April 2010; received in revised form 15 April 2011; accepted 15 April 2011

Abstract

This is a prospective cohort study to identify factors associated with receipt of substance abuse treatment (SAT) among adults with alcohol problems and HIV/AIDS. Data from the HIV Longitudinal Interrelationships of Viruses and Ethanol study were analyzed. Generalized estimating equation logistic regression models were fit to identify factors associated with any service utilization. An alcohol dependence diagnosis had a negative association with SAT (adjusted odds ratio [AOR] = 0.36, 95% confidence interval [95% CI] = 0.19–0.67), as did identifying sexual orientation other than heterosexual (AOR = 0.46, CI = 0.29–0.72) and having social supports that use alcohol/drugs (AOR = 0.62, CI = 0.45–0.83). Positive associations with SAT include presence of hepatitis C antibody (AOR = 3.37, CI = 2.24–5.06), physical or sexual abuse (AOR = 2.12, CI = 1.22–3.69), social supports that help with sobriety (AOR = 1.92, CI = 1.28–2.87), homelessness (AOR = 2.40, CI = 1.60–3.62), drug dependence diagnosis (AOR = 2.64, CI = 1.88–3.70), and clinically important depressive symptoms (AOR = 1.52, CI = 1.08–2.15). While reassuring that factors indicating need for SAT among people with HIV and alcohol problems (e.g., drug dependence) are associated with receipt, nonneed factors (e.g., sexual orientation, age) that should not decrease likelihood of receipt of treatment were identified.

Keywords: Substance abuse; Treatment; Addiction; HIV/AIDS; Alcohol

1. Introduction

An estimated 44% of adults living with HIV/AIDS have alcohol and other drug (AOD) use disorders, which is much higher than the general population (Bing et al., 2001; Galvan et al., 2002; Galvan, Burnam, & Bing, 2003; Rabkin, Ferrando, Jacobsberg, & Fishman, 1997; Rabkin, McElhiney, & Ferrando, 2004). AOD use by people with HIV has particularly deleterious results, including an increased burden on support and medical systems (Masson, Sorensen, Phibbs, & Okin, 2004), a higher likelihood of engagement in risk behaviors that result in infection transmission (Palepu et al., 2005), nonadherence with HIV/AIDS treatments and faster biomedical decline (Conigliaro et al., 2004; Dausey & Desai, 2003; Kelley & Petry, 2000; Lucas, Gebo, Chaisson, & Moore, 2002; Meyerhoff, 2001; Petry, 1999), and liver complications, particularly considering the impact of co-occurring alcohol use, HIV/AIDS medications, and hepatitis C (Lucas et al., 2002; Moore, Keruly, & Chaisson, 2004; Palepu et al., 2003; Petry, 1999; Samet et al., 2007). Despite the magnitude of

* Corresponding author. Loyola University of Chicago School of Social Work, 820 N. Michigan Avenue, 12th floor, Chicago, IL 60611 USA. Tel.: +1 312 915 7494; fax: +1 312 915 7645.
E-mail address: jorwat@luc.edu (J. Orwat).

0740-5472/11/$ – see front matter
doi:10.1016/j.jsat.2011.04.002

substance abuse problems among people living with HIV, individual factors other than need influence receipt of SAT for those who might benefit (Burnam et al., 2001).

Substance abuse treatment (SAT) can be effective in reducing these consequences (Klinkenberg & Sacks, 2004; Loughlin et al., 2004; Palepu, Horton, Tibbetts, Meli, & Samet, 2004), although outcomes are more complex than mere attendance. These include characteristics of the treatment system, such as staff characteristics and availability (Hser, 1995; Najavits & Weiss, 1994; Najavits, Crits-Christoph, & Dierberger, 2000) and the relationship with the client outside of what occurs within treatment sessions (Najavits & Weiss, 1994). SAT outcomes are also dependent on the severity of other significant life problems, such as family relations, employment, or medical and mental health problems (McLellan & Weisner, 1996), which necessitates matching clients to specific treatment services that best treat co-occurring problems (Gastfriend & McLellan, 1997). Ultimately, the interaction of many interpersonal and intrapersonal factors may impact the potential for relapse or chronic relapse of a client utilizing SAT (Marlatt, 1985, 1996a, 1996b; Marlatt, Barrett, & Daley, 1999; McLellan & McKay, 2003).

Burnam et al. (2001) analyzed the HIV Cost and Services Utilization Study data, a national probability survey of adults with HIV receiving medical care in 1996. Controlling for need, outpatient AOD service utilization was positively associated with lower income levels and lower educational levels. African Americans were more likely to access AOD treatment services. Respondents who identified as gay were less likely to utilize AOD services but were more likely to seek treatment through medical providers and use psychiatric medications. Insurance coverage had little impact on utilization. Those who are not employed and those with disability were more likely to use outpatient treatment. Only one indicator of HIV severity, lower CD4 cell count, was positively associated with AOD services.

The identification of individual-level factors is particularly important because HIV/AIDS disproportionately strikes vulnerable populations who may experience barriers, such as people living in poverty and racial/ethnic minorities, particularly African American women, gay men, and other disenfranchised populations (Orwat, 2004). This exploratory analysis adds to the literature by examining the factors associated with SAT in the context of contemporary HIV-related medical practice including highly active antiretroviral therapies and examined factors associated with SAT utilization in a prospective cohort of individuals living with HIV/AIDS and AOD problems. Secondary analyses of the factors associated with four specific types of treatment (detoxification program, residential, outpatient, and "other" treatment) were also conducted. Andersen's socio-behavioral model, as adapted for vulnerable populations, guided the identification and organization of factors of interest (Andersen & Newman, 1973; Aday & Andersen, 1974; Andresen, Malmgren, et al., 1994).

2. Materials and methods

Data from the HIV Longitudinal Interrelationships of Viruses and Ethanol (HIV-LIVE) study were analyzed, a prospective cohort study of 400 adults with HIV infection and alcohol problems. Eligibility criteria included documentation of HIV infection, a current or lifetime alcohol problem as evidenced by two or more positive responses to the CAGE (Cut down, Annoyed, Guilty, and Eye opener) alcohol screening questionnaire (Ewing, 1984; Samet, Phillips, Horton, Traphagen, & Freedberg, 2004) or an alcohol use disorder in the judgment of a physician investigator, an ability to speak English or Spanish, a score of 21 or greater on the Mini Mental State Examination (Folstein, Folstein, & McHugh, 1975; Smith, Horton, Saitz, & Samet, 2006), and plans to remain in the area for at least 1 year. Classification of an alcohol problem ("none," "moderate," "at risk") was derived from the National Institute on Alcohol Abuse and Alcoholism's (NIAAA's) definition of amounts that risk consequences (Samet et al., 2007). The current analysis included subjects with at least two study interviews. All subjects were 18 years or older and provided written informed consent. A certificate of confidentiality was provided by NIAAA as an additional protection of subject privacy.

Subjects were recruited from two urban hospital-based outpatient practices (an intake clinic for HIV-positive clients at Boston Medical Center and a specialty clinic for HIV primary care at Beth Israel Deaconess Medical Center in Boston) by referral of a practicing physician and from flyers posted in the community, medical clinics, and SAT facilities. Approximately 38% of participants were recruited from previous observational studies conducted by these investigators. The largest proportion of subjects was recruited from the Boston Medical Center (43%), and all subjects resided in the Boston area.

After informed consent was obtained, subjects were enrolled from August 2001 to July 2003. Data were collected at baseline and every 6 months thereafter for up to 42 months. Data collection ended in March 2006. Domains of data collected included demographics, HIV risk behaviors, AOD severity and diagnosis, health care utilization, SAT service utilization, indicators of HIV disease progression, and medical comorbidities. Extensive efforts were used to maximize follow-up: 100% had at least one follow-up interview (two total interviews) with a median of three follow-up interviews (four total); the 25th percentile was two follow-up interviews (three total), and the 75th percentile was five follow-ups (six total interviews).

2.1. Dependent variables

The primary outcome was self-reported receipt of any SAT in the past 6 months, which was assessed by asking subjects, "During the past 6 months, did you receive any substance abuse services?" For the four secondary outcomes of specific types of SAT (i.e., detoxification program,

residential, outpatient, and other treatment), self-reported utilization was evaluated by asking, "How many days did you receive substance abuse services from each of the following programs during the past 6 months?" Subjects were asked about the following specific types of SAT: a detoxification program, a halfway house or residential facility, a substance abuse counselor in an outpatient program, a day treatment program, and/or some other type of treatment, such as a doctor, a priest or a rabbi, or an employee assistance program. Because of low rates of participation, day treatment was excluded from multivariable analysis (Table 2). Attendance at Alcoholics Anonymous and Narcotics Anonymous Groups was not examined in this analysis and was reported separately (Orwat et al., 2011).

2.2. Independent variables

The Andersen model was used to identify independent variables potentially associated with SAT and organize them into predisposing characteristics, enabling resources, and need variables (Andersen, 1995; Weisner, Matzger, Tam, & Schmidt, 2002). This model is commonly used in studies to understand utilization of health care by examining three individual-level domains: those factors that predispose people to utilize care (e.g., demographics), factors that enable use (e.g., ability to pay), and factors associated with need for the specific service (Andersen, 1995). The rich database allowed for the exploration of other factors, which have yet to be tested in the literature (e.g., literacy, hepatitis C).

Predisposing characteristics consisted of age (defined as greater than the median, 44.5), gender, race (Black, White, and other; because of small sample sizes, other includes other race and Hispanic), marital status (single, married, or partnered but not married), sexual orientation (gay, bisexual, other vs. heterosexual), education level (high school education), living alone, born in the United States, sexual or physical abuse or trauma, and high school literacy level (defined as a score >60 on the Rapid Estimate of Adult Literacy in Medicine (Davis et al., 1991). HIV/AIDS disease severity was considered predisposing variable to remain consistent with the literature (Burnam et al., 2001) and was assessed using CD4 cell count (greater than vs. less than or equal to 350 cells/mm^3, the median, CD4 count), the presence of opportunistic infections or cancers, and the HIV symptom index score—a measure of how often and bothersome clients experience 20 common HIV symptoms (Justice et al., 2001; Kilbourne et al., 2002). The presence or absence of hepatitis C virus (HCV) antibody indicated exposure. Testing for HCV was completed at the first opportunity if the results of HCV testing for clinical purposes were not available.

Enabling resources included employment status, social support use of alcohol or drugs, social support that is supportive of sobriety, living with children, receipt of government disability income, and homelessness (spending one or more nights "on the street, without shelter" over the past 6 months). Social support use of alcohol or drugs was assessed by asking, "how many of the people you spend time with," followed by three prompts, "currently drink alcohol," "are currently heavy or problem drinkers," and "currently use drugs." Social support for sobriety was assessed by asking, "How many of the people that you spend time with support your sobriety or abstinence." The use of HIV antiretroviral medications was assessed by asking about specific medication use over the previous 6 months as listed with photographic prompts and generic and brand names. Cohort members were not assessed with regard to use of naltrexone or other medications used to treat substance use disorder at the time of the study; its use for alcohol problems, although indicated, occurred rarely in practice.

Need/Severity for SAT included alcohol dependence in the past 6 months, which was assessed using the Composite International Diagnostic Interview Short Form (CIDI-SF; Kessler, Andrews, Mroczeek, et al., 1998), and the diagnosis of drug dependence in the previous 6 months. Depressive symptoms were assessed using the Center for Epidemiologic Studies Depression scale (CES-D; Andresen, Malmgren, Carter, & Patrick, 1994; Radloff, 1977) and dichotomized at a score of 16 or greater, which indicates clinically important depressive symptoms in the general population (Eaton & Kessler, 1981; Weissman, Sholomskas, Pottenger, Prusoff, & Locke, 1977). Involvement in the criminal justice system was measured by asking respondents if they were in jail during the previous 6 months.

2.3. Analysis

To account for correlation in the data due to incorporating repeated measures from the same subject, the multivariable analysis that explored the association between selected factors and SAT utilization was conducted using generalized estimating equations (GEE; Zeger & Liang, 1986; Liang & Zeger, 1986) logistic regression models. Models were fit using a logit link, an exchangeable working correlation structure, and standard errors were based on the empirical sandwich estimator for all analyses. The GEE approach yields robust results (i.e., valid estimates can be obtained from the empirical variance estimator even if the correlation structure is misspecified), is flexible, and can accommodate analyses where subjects have varying numbers of follow-up interviews. The following predisposing characteristics were modeled as time-dependent variables: married, partnered, living alone, and all HIV status variables; all enabling resources and need/severity variables were modeled as time-dependent variables. Time-varying independent variables were "lagged" to predict SAT utilization in the subsequent interview to ensure such factors preceded the episode of SAT. If a participant missed a follow-up interview, the outcome at the next available interview was used; the models adjusted for the duration of time between interviews for the

lagged independent variables and outcomes. In addition, the time that the outcome was assessed (number of months since baseline) was included as a covariate to account for potential temporal trends in the data. Preliminary crude analyses were performed that included only a single independent variable. A full multivariable model was then fit including all independent variables in the same model. Potential collinearity was initially assessed by verifying that no pair of independent variables was highly correlated (>0.40). In addition, the variance inflation factor (VIF) was calculated for each independent variable and covariate to further evaluate potential collinearity. A criterion of VIF greater than 10 was used to indicate collinearity (Kutner, Nachtsheim, & Neter, 2004). The largest observed VIF was 1.84, suggesting multicollinearity was likely not an issue in our regression models.

Analyses of the four secondary outcomes were performed using the same approach as described for the primary outcome. Analyses were conducted using two-sided tests and a significance level of .05. Because of the exploratory nature of the analyses, we did not adjust for multiple comparisons. All analyses were conducted using SAS version 9.0 (SAS Institute, Inc., Cary, NC).

3. Results

3.1. Sample characteristics

Of the 400 HIV-LIVE subjects, 369 completed at least two research interviews and therefore constituted the study sample. Baseline characteristics are displayed in Table 1. Most subjects were non-White and male, and the mean age was 42.7 years. A small minority were gay or bisexual, whereas a vast majority had experienced sexual or physical trauma or abuse. Few had an opportunistic condition, mean CD4 count was 462 cells/mm^3, and more than half had a positive hepatitis C antibody test (58%). Most reported spending time with people who use alcohol, whereas less than half spent time with people who use drugs; a majority had social support for sobriety. At baseline, 31.5% were classified as "at-risk" drinkers by the NIAAA classification for an alcohol problem, 10.8% as "moderate" drinkers, and 57.8% as nondrinkers. These numbers varied somewhat over time and across observations (Bertholet, Cheng, Samet, Quinn, & Saitz, 2010). Criteria for a current (past 6 months) alcohol dependence diagnosis were met by 12% of the sample, and 43% met the criteria for a current (past 6 months) drug dependence diagnosis. At study entry, almost half (45%) had used SAT services in the preceding 6 months, 18% detoxification services, 20% residential, 17% outpatient counseling, and 6% some other treatment (Table 1). In subsequent interviews, 40% (n = 133) utilized SAT in the 6 months prior to the second interview, 34% (n = 94) prior to the third, 35% (n = 64) prior to the fourth, 30% (n = 35) prior to the fifth, 28% (n = 14) prior to the sixth, and 33% (n = 1) prior to the eighth.

Table 1
Baseline characteristics of 400 subjects recruited for HIV-LIVE study (N = 369)

Variable	Study sample
SAT utilization [a]	
Any SAT services [b]	167 (45.3)
Detoxification [c]	65 (17.6)
Residential	74 (20.1)
Outpatient counseling	63 (17.1)
Other treatment [d]	22 (6.0)
Predisposing characteristics: traditional domains	
Age, years, *M* (*SD*)	42.7 (7.5)
Gender (female)	92 (24.9)
Race	
White	120 (32.5)
Black	158 (42.8)
Other	91 (24.7)
Married	24 (6.7)
High school graduate [a]	240 (65.0)
Lives alone	106 (28.7)
Predisposing characteristics: vulnerable domains	
Born in the United States	323 (87.5)
Literacy, high school level [e]	235 (63.7)
Gay, bisexual, other sexual orientation	121 (32.8)
Any traumatic abuse, ever [f]	297 (80.4)
HIV severity variables	
Any opportunistic conditions [g]	61 (16.6)
HIV quality of life scale, *M* (*SD*) [h]	9.6 (4.9)
CD4 count (cells/mm^3), *M* (*SD*)	462.1 (299.4)
HCV antibody present	214 (58.0)
Enabling resources: traditional domain	
Employment status (unemployed) [g]	271 (73.4)
Social support uses alcohol or drugs	
Social support drinks alcohol	228 (61.7)
Social support uses drugs	174 (47.3)
Social support helps with sobriety	327 (88.8)
Currently taking anti-HIV medications	235 (63.7)
Enabling resources: vulnerable domain	
Received disability [g]	271 (73.4)
Homeless [g]	89 (24.1)
Need/Severity	
Alcohol diagnosis [g,i]	43 (11.8)
Drug dependence diagnosis [g]	159 (43.1)
Depressive symptoms [j]	230 (62.3)
In jail [g]	67 (18.2)

Note. Values are number (percentage) unless otherwise indicated.

[a] High school graduate (12 or more years of education).

[b] Any SAT was assessed by "…did you receive any substance abuse treatment services?"

[c] *Detoxificaion* is defined as any in- our outpatient program.

[d] Other treatment such as a doctor, a priest or a rabbi, or an employee assistance program.

[e] Literacy score higher than 60 (Davis et al., 1991).

[f] Sexual or physical trauma or abuse.

[g] In the past 6 months.

[h] Justice et al. (2001) and Kilbourne et al. (2002).

[i] Alcohol diagnosis assessed using CIDI-SF (Kessler et al., 1998).

[j] Measure of depressive symptoms where CES-D score 16 or higher indicates substantial depressive symptoms.

3.2. Multivariable models: Predictors of any SAT

Predisposing characteristics positively associated with any SAT included physical or sexual trauma or abuse

and the presence of HCV antibody (Table 2). Identifying as gay, bisexual, or other (compared with heterosexual) and CD4 cell count were negatively associated with SAT. Enabling resources positively associated included social support help with sobriety, living without children, and homelessness; a social support system using alcohol and/or drugs was negatively associated. Alcohol dependence was negatively associated with SAT; a drug dependence diagnosis and depressive symptoms were positively associated.

3.3. Multivariable models: Predictors of specific types of SAT

3.3.1. Detoxification

Older age and lower CD4 count were associated with lower odds of participation in a detoxification program, whereas the presence of HCV antibody was positively associated (Table 3). As for enabling resources, homelessness was positively associated with detoxification, as were the need factors of drug dependence and depressive symptoms.

Table 2
Factors associated with any SAT[a]

Variable	Unadjusted model Odds ratio (95% CI)	Fully adjusted model (n = 1,153) Odds ratio (95% CI)
Predisposing characteristics: traditional domains		
Age[b]	1.10 (0.88–1.38)	0.98 (0.95–1.01)
Gender (female)	1.74 (1.35–2.25)***	1.24 (0.75–2.03)
Race		
Black	1.03 (0.80–1.32)	1.16 (0.72–1.89)
Other	1.40 (1.03–1.89*	0.85 (0.45–1.61)
Married	1.49 (0.99–2.31)	1.67 (0.83–3.34)
Partnered, not married	1.33 (1.06–1.68)*	1.35 (0.95–1.91)
High school graduate[c]	0.70 (0.55–0.88)**	0.96 (0.62–1.47)
Predisposing characteristics: vulnerable domains		
Lives alone	0.47 (0.37–0.61)***	0.75 (0.54–1.05)
Born in the United States	0.74 (0.52–1.03)	0.62 (0.3–1.29)
Literacy, high school level[d]	0.93 (0.73–1.18)	1.27 (0.82–1.95)
Gay, bisexual, other sexual orientation	0.29 (0.22–0.38)***	0.46 (0.29–0.72)***
Any traumatic abuse, ever[e]	1.57 (1.18–2.09)**	2.12 (1.22–3.69)
HIV status variables		
Any opportunistic conditions[f]	1.29 (0.99–1.69)	0.72 (0.50–1.04)
HIV quality of life scale[g]	1.20 (1.59–2.51)***	1.01 (0.97–1.05)
CD4 count[h]	0.86 (0.68–1.08)	0.95 (0.90–1.00)*
HCV antibody present	4.12 (3.21–5.25)***	3.37 (2.24–5.06)***
Enabling resources: traditional domain		
Unemployed[f]	2.82 (2.13–3.73)***	1.29 (0.88–1.9)
Social support uses alcohol or drugs	0.58 (0.45–0.74)***	0.62 (0.45–0.83)**
Social support helps with sobriety	2.08 (1.42–3.04)***	1.92 (1.28–2.87)**
Lives without kids	1.04 (0.71–1.54)	1.89 (1.09–3.27)*
Currently taking anti-HIV medications	0.67 (0.53–0.84)***	0.96 (0.68–1.36)
Enabling resources: vulnerable domain		
Received disability[f]	2.11 (1.56–2.86)***	0.87 (0.52–1.46)
Homeless[f]	3.25 (2.42–4.36)***	2.4 (1.6–3.62)***
Need/Severity		
Alcohol diagnosis[f,i]	0.84 (0.59–1.19)	0.36 (0.19–0.67)**
Drug dependence diagnosis	4.04 (3.14–5.19)***	2.64 (1.88–3.70)***
Depressive symptoms[j]	2.88 (2.26–3.68)***	1.52 (1.08–2.15)*
In jail[d]	2.34 (1.64–3.35)***	0.99 (0.58–1.70)

* $p < .05$.
** $p < .01$.
*** $p < .001$.
[a] Unit of analysis is person-observation period. These are observations not patients.
[b] Age is a dichotomous variable at median (44.5 years).
[c] High school graduate is 12 or more years of education)
[d] Literacy score higher than 60 (Davis et al., 1991).
[e] History of sexual or physical traumatic abuse.
[f] Past 6 months; time varying.
[g] Justice et al. (2001) and Kilbourne et al. (2002).
[h] CD4 count a dichotomous variable at median (407.5 cells/mm^3).
[i] Alcohol diagnosis using CIDI (Kessler et al., 1998).
[j] Measure of depressive symptoms where a CES-D score 16 or higher indicate depressive symptoms.

Table 3
Factors associated with types of SAT[a]

Variable	Detox (n = 1,153) AOR (95% CI)	Residential (n = 1,153) AOR (95% CI)	Outpatient (n = 1,153) AOR (95% CI)	Other treatment (n = 1,153) AOR (95% CI)
Predisposing characteristics: traditional domain				
Age[b]	0.96 (0.92–0.99)*	0.98 (0.95–1.02)	1.01 (0.98–1.04)	0.97 (0.93–1.02)
Gender (female)	0.74 (0.41–1.33)	0.77 (0.42–1.41)	1.59 (0.93–2.71)	0.84 (0.38–1.90)
Race				
Black	1.56 (0.94–2.59)	1.7 (0.93–3.09)	1.61 (0.95–2.72)	1.59 (0.78–3.24)
Other	0.73 (0.34–1.57)	0.49 (0.2–1.2)	1.02 (0.46–2.25)	2.45 (1.03–5.83)*
Married	1.06 (0.45–2.46)	0.45 (0.15–1.39)	1.13 (0.54–2.38)	1.81 (0.67–4.88)
Partnered, not married	1.02 (0.62–1.68)	0.63 (0.38–1.03)	1.36 (0.94–1.97)	1.33 (0.76–2.35)
High school graduate[c]	1.18 (0.76–1.84)	1.16 (0.67–2.02)	1.41 (0.83–2.42)	0.97 (0.53–1.78)
Predisposing characteristics: vulnerable domain				
Lives alone	0.99 (0.58–1.68)	0.48 (0.27–0.86)	0.93 (0.61–1.4)	0.82 (0.40–1.70)
Born in the United States	0.86 (0.35–2.09)	0.57 (0.21–1.55)	0.55 (0.21–1.42)	1.90 (0.61–5.97)
Literacy, high school level[d]	1.14 (0.7–1.87)	1.18 (0.67–2.1)	1.32 (0.8–2.2)	1.27 (0.62–2.6)
Gay, bisexual, other sexual orientation	0.63 (0.37–1.09)	0.59 (0.32–1.11)	0.86 (0.52–1.44)	0.82 (0.34–1.98)
Any traumatic abuse, ever[e]	1.79 (0.88–3.61)	1.78 (0.9–3.49)	1.47 (0.77–2.81)	2.13 (0.94–4.86)
HIV status variables				
Any opportunistic conditions[f]	1.29 (0.79–2.09)	0.65 (0.39–1.07)	0.77 (0.5–1.17)	0.56 (0.25–1.28)
HIV quality of life scale[g]	0.98 (0.93–1.03)	1.03 (0.98–1.08)	1.02 (0.97–1.07)	1 (0.93–1.08)
CD4 count[h]	0.91 (0.84–0.99)*	0.99 (0.92–1.06)	1 (0.93–1.07)	1 (0.91–1.11)
HCV antibody present	2.02 (1.22–3.33)**	2.26 (1.27–4)	3.2 (1.88–5.43)***	1.18 (0.64–2.17)
Vulnerable domain: traditional domain				
Unemployed[f]	0.8 (0.45–1.41)	1.28 (0.71–2.3)	1.16 (0.69–1.96)	2.55 (1.09–5.96)*
Social support uses alcohol or drugs	1.38 (0.87–2.19)	0.46 (0.29–0.71)	0.57 (0.39–0.85)**	0.86 (0.5–1.49)
Social support helps with sobriety	1.1 (0.61–2.01)	2.15 (1.05–4.41)	1.58 (0.85–2.96)	5.21 (1.34–20.32)*
Lives without kids	1.63 (0.62–4.24)	15.58 (2.13–114.16)	2.22 (1.07–4.61)*	1.19 (0.4–3.55)
Currently taking anti-HIV medications	0.75 (0.48–1.16)	0.98 (0.61–1.58)	0.93 (0.61–1.41)	0.7 (0.38–1.26)
Enabling resources: vulnerable domain				
Received disability[f]	1.31 (0.65–2.64)	0.7 (0.36–1.4)	1.22 (0.63–2.34)	1.64 (0.64–4.18)
Homeless[f]	3.29 (2.00–5.41)***	2.05 (1.26–3.33)	1.84 (1.15–2.97)*	1.71 (0.93–3.15)
Need/Severity				
Alcohol diagnosis[f,i]	0.78 (0.38–1.58)	1.2 (0.58–2.46)	0.49 (0.24–1)	0.63 (0.23–1.73)
Drug dependence diagnosis	3.31 (2.00–5.49)***	2.26 (1.42–3.59)	1.7 (1.13–2.57)*	1.35 (0.78–2.33)
Depressive symptoms[j]	2.76 (1.58–4.82)***	1.35 (0.72–2.52)	1.56 (0.96–2.54)	0.94 (0.51–1.74)
In jail[d]	1.5 (0.76–2.93)	1.5 (0.8–2.81)	1.13 (0.61–2.08)	0.94 (0.44–2.03)

Note. AOR = adjusted odds ratio.

* $p < .05$.
** $p < .01$.
*** $p < .001$.
[a] Unit of analysis is person-observation period. These are observations not patients.
[b] Age is a dichotomous variable at median (44.5 years).
[c] High school graduate is 12 or more years of education)
[d] Literacy score higher than 60 (Davis et al., 1991).
[e] History of sexual or physical traumatic abuse.
[f] Past 6 months; time varying.
[g] Justice et al. (2001) and Kilbourne et al. (2002).
[h] CD4 count a dichotomous variable at median (407.5 cells/mm^3).
[i] Alcohol diagnosis using CIDI (Kessler et al., 1998).
[j] Measure of depressive symptoms where a CES-D score 16 or higher indicate depressive symptoms.

3.3.2. Residential treatment

Of the predisposing characteristics, living alone was negatively associated with residential treatment and the presence of HCV antibody was positively associated, as was the need factor of a drug dependence diagnosis, but no HIV disease severity variable was significantly associated with receipt of residential treatment (Table 3). Enabling resources positively associated with residential treatment included social support for sobriety, living without children, and homelessness; social support use of alcohol/drugs was negatively associated.

3.3.3. Outpatient treatment

The presence of HCV antibody was positively associated with outpatient treatment, the only statistically significant predisposing characteristic (Table 3). Living without children and homelessness were also positively associated, whereas social support use of alcohol and drugs was

negatively associated. A drug dependence diagnosis was associated with higher odds of outpatient treatment.

4. Discussion

This exploratory study of those living with HIV/AIDS and alcohol or drug problems identified a number of nonneed factors associated with receipt of SAT. These associations were found for any SAT and for each of four types of SAT: detoxification program, residential treatment, and outpatient treatment. Identification of the factors that facilitate or impede use of effective SAT suggests areas in need of improvement. Specifically, we identified a negative association between gay, lesbian, or other sexual orientation and SAT utilization, whereas past physical or sexual abuse or trauma was positively related. We detected a negative association between median CD4 cell count and utilization of SAT, although the presence of the HCV antibody was strongly and positively associated. Social supports were important factors related to receipt of SAT; supports that help with sobriety were positively associated, and those that used alcohol or drugs, negatively associated. Homelessness had a positive and strong association with SAT. In this sample of people with alcohol problems, most of whom had lifetime dependence, current alcohol dependence diagnosis was negatively associated with SAT, although a current drug dependence diagnosis and depressive symptoms were each positively associated.

This research partially supports studies that demonstrate SAT is underutilized by sexual minorities in the general population (Goldstein et al., 2005) and those living with HIV/AIDS (Burnam et al., 2001; Goldstein et al., 2005). Burnam et al. (2001) found that among those with HIV/AIDS, those who identify as gay were less likely to participate in residential or outpatient treatment. This study examined the effect of sexual orientation among people living with HIV/AIDS by controlling for variables such as homelessness, sexual or physical abuse or trauma, and social support for substance use or abstinence. Lower SAT utilization may be related to concerns about the following: the treatment setting (e.g., emotional and physical safety) if sexual orientation is disclosed; insensitivity to the experience of antigay discrimination; lack of appreciation for nontraditional support systems; attempts to change the person's sexual orientation; and/or the failure to understand nontraditional forms of socialization (Cochran & Cauce, 2006).

We did not detect an independent association of race and gender on utilization of treatment, which is somewhat inconsistent with previous studies that suggest that women are less likely than men to utilize treatment and African Americans are more likely than Whites to utilize treatment (Burnam et al., 2001; Goldstein et al., 2005; Greenfield et al., 2007). This may be due to regional differences, such as Boston's diverse SAT resources, which may have increased efforts to reach minorities and women as well as to design appropriate treatment programs (McAuliffe & Dunn, 2004). However, this important issue warrants further exploration, especially in light of the increasing prevalence of HIV/AIDS in ethnic and racial minorities and women.

The positive association of past sexual or physical abuse or trauma and SAT is congruent with the literature, although this analysis supported the association in a sample living with HIV/AIDS and controlling for factors often related to an increased likelihood of trauma, such as being female, gay, lesbian, or bisexual and race (Farley, Golding, Young, Mulligan, & Minkoff, 2004). Close to 90% of individuals seeking treatment for substance use disorders have had an experience of trauma (Farley et al., 2004). Individuals exposed to trauma are more likely to experience a wide range of negative outcomes that complicate SAT, including an increased likelihood of mental health problems, such as posttraumatic stress disorder (Kayo & Rojas, 2004), which increase the likelihood of relapse (Farley et al., 2004; Wadsworth, Stampneto, & Halbrook, 1995). Previous treatment experiences for these individuals may not have adequately addressed issues related to trauma (Farley et al., 2004).

Few measures of HIV disease severity were associated with treatment, although our findings suggested that people with more advanced HIV are more likely to receive SAT. This is consistent with previous studies (Burnam et al., 2001). The strong and positive association of hepatitis C and SAT may be the result of clinical and/or policy efforts to facilitate access to treatment or by patient concerns. Tsui et al. (2007) observed that people who knew their hepatitis C status were more likely to abstain from alcohol. We speculated that it may also be the case that when concern rises about hepatitis C, people make attempts to decrease substance use and seek SAT. Therapies for hepatitis C are complex, requiring weekly injections that have adverse effects and hold no guarantees of cure. With increased treatment complexities related to patients with AOD problems with co-occurring hepatitis C and HIV, clinicians may influence such patients being treated for hepatitis C to also address substance abuse issues (Mehta et al., 2005). However, most people in this cohort had barriers to treatment of hepatitis C (Nunes et al., 2006). Therefore, the association between hepatitis C and SAT may also be explained by testing occurring in or by referral from SAT programs. Ultimately, a multidisciplinary approach to SAT and co-occurring HIV and hepatitis C that includes pharmacotherapy, mental health services, along with medical care may be most effective (Nunes et al., 2006).

The social support findings in this study were consistent with the general population studies (Hasin & Grant, 1995; Kaskutas, Bond, & Humphreys, 2002; Witbrodt & Kaskutas, 2005). This analysis found that social support matters, as social support that is supportive of sobriety influences utilization of services, whereas social supports that are currently using substances, do not. Not only does research show the importance of social support in the utilization of

substance abuse services, but it also affirms the role of social support in HIV treatment adherence (Knowlton et al., 2006).

The significant positive association of homelessness with SAT utilization is consistent with Goldstein et al.'s (2005) analysis of people living with HIV/AIDS; the current analysis has similar findings when controlling for factors often found to be related to SAT among the homeless, such as gender, race/ethnicity, and social support (Kertesz et al., 2006; Song, 1999). The association of homelessness with detoxification services was particularly strong, which may be related to triage decisions in which few options exist for emergency housing, particularly for those that are currently using substances (Kertesz et al., 2006; Song, 1999). However, service utilization may also be the result of residing in temporary shelters or participation in long-term housing programs, which may require SAT as part of the programmatic requirements or as a condition of participation (Kertesz et al., 2006).

We found a significant positive association of a drug dependence diagnosis and depressive symptoms on the receipt of SAT, but a current alcohol dependence diagnosis was negatively associated. This finding may seem curious at first glance but may be related to the sample, almost all of whom likely have lifetime alcohol dependence and many have co-occurring drug dependence. Drug dependence may lead to seeking specific care (e.g., pharmacotherapy), whereas those with current alcohol dependence are less likely to do so.

This study had notable strengths and limitations. First, this analysis identified association, not causation, between various factors and SAT. For example, it may be that treatment led to social support rather than social support motivating SAT. We attempted to capture the temporal relationship between variables by "lagging" outcome variables so that they were taken from the period subsequent to explanatory factors. A second limitation may be the reliance on self-report of SAT utilization, although asking whether SAT occurred is valid and reliable, particularly when not seeking quantity of treatment (Goldberg, Seybolt, & Lehman, 2002). However, self-report of SAT may capture interventions that may not appear in comprehensive databases (e.g., from a personal physician or clergy member).

Because of concerns related to sample size, we do not report a subanalysis of those individuals with the highest need (i.e., meeting 6-month criteria for alcohol and/or drug dependence at follow-up time points), although that might have provided further insight into SAT utilization among those with the highest need. Additional limitations may relate to generalizability of the findings based on the recruitment site(s), urban setting with a spectrum of sexual orientation representation, many homeless, and minorities. This acknowledged limitation is also a strength given the epidemiology of this epidemic that disproportionately affects this population. Bias in self-report may have been attenuated in this study by use of validated survey questions and the minimization of contextual factors that may influence the possibility of biased self-report (Del Boca & Darkes, 2003).

Identifying as gay, lesbian, or some other sexual orientation was negatively associated with SAT; however, the association was not significant when specific types of SAT were examined. Although previous research supports this finding, additional analysis with sexual minorities is to further examine the interaction of sexual orientation with other factors that may be correlated, such as marital status or living with children.

Future analysis might include some variables not available in this analysis to address potential limitations of this study. One specific example is the lack of a measure concerning respondent health insurance status and benefit structure. People living with HIV in Massachusetts have access to health insurance (Orwat, 2004), so despite the absence of this variable, we assume all participants had access. Future studies may examine the impact of different insurance products (i.e., public, private, and none) and benefit structures (i.e., networks and copayments) on SAT utilization. Other variables potentially associated with SAT that were not included in this analysis included lifetime experience with SAT attendance at more than one type of SAT concurrently, the focus of treatment on admission (e.g., alcohol, drugs, substance abuse/mental health), clinical orientation of treatment programs utilized, or use of addiction pharmacotherapy (e.g., naltrexone).

One strength of the study is the clinical detail, which enabled consideration of variables not previously explored, including social support, HIV/AIDS disease severity, and hepatitis C. A second strength is the prospective research quality of assessments and follow-ups. Third, the analysis adjusted for several potential confounding variables and used lagged analyses to better capture the temporal relationship between the independent and dependent variables. Finally, the study used a strong theory-based approach in the use of the Andersen model.

This research provides a better understanding of the factors that facilitate or impede SAT for HIV-infected persons, particularly factors not related to treatment need, which informs the development of strategies to increase utilization. Barriers to treatment include having a sexual orientation other than heterosexual and social supports that use alcohol and/or drugs. Sexual or physical abuse or trauma, the presence of HCV antibody, a social network that supports sobriety, homelessness, drug dependence, and depressive symptoms were factors with a positive association with SAT. Removing barriers to SAT for affected HIV-infected persons includes the consideration of systemic and individual-level factors. Individuals can be helped to seek social networks that support sobriety. System-level interventions may include the development of specialty interventions for populations at risk, such programs for those who are not heterosexual. The findings of this study may assist clinicians as they motivate people living with HIV/AIDS to utilize SAT and administrators as they consider system design, particularly for those who access both SAT and medical services.

Acknowledgments

Support for this study came from the following grants from the NIAAA of the NIH: R01-AA13216 (Clinical Impact of HCV and Alcohol in HIV-Infected Persons), R01-AA11785 (Medication Adherence in Alcohol Abusing HIV Patients), R01-AA10870 (Enhanced Linkage of Alcohol Abusers to Primary Care), K24-AA015674 (Impact of Alcohol Use on HIV Infection—in the United States and Russia), and the Ruth L. Kirschstein National Research Service Award for Individual Predoctoral Fellows (National Institute on Drug Abuse—F31 DA021118 [Predictors of SA Treatment among Adults Living with HIV]). This research was conducted in part in the General Clinical Research Center at Boston University School of Medicine (USPHS Grant M01 RR00533) and the Beth Israel Deaconess General Clinical Research Center (USPHS Grant M01 RR01032).

The authors thank Carly Bridden for project support and organization; DM-STAT for meticulous data management; Tamara Faulkner for editorial and research assistance; the Clinical Addiction Research and Education Unit Staff Research Associates and Project Managers for the difficult tasks of enrollment, follow-up, and quality of the data collected; and the subjects for their contribution of time and effort.

References

Aday, L., & Andersen, R. (1974). A framework for the study of access to medical care. *Health Services Research*, *9*, 208–220.

Andersen, R. M. (1995). Revisiting the behavioral-model and access to medical care—Does it matter. *Journal of Health and Social Behavior*, *36*, 1–10.

Andersen, R., & Newman, J. F. (1973). Societal and individual determinants of medical care utilization in the United States. *Milbank Memorial Fund Quarterly*, *51*, 95–124.

Andresen, E. M., Malmgren, J. A., Carter, W. B., & Patrick, D. L. (1994). Screening for depression in well older adults—Evaluation of a short-form of the CES-D. *American Journal of Preventive Medicine*, *10*, 77–84.

Bertholet, N., Cheng, D. M., Samet, J. H., Quinn, E., & Saitz, R. (2010). Alcohol consumption patterns in HIV-infected adults with alcohol problems. *Drug Alcohol Depend*, *112*, 160–163.

Bing, E. G., Burnam, A., Longshore, D., Fleishman, J. A., Sherbourne, C. D., London, A. S., et al. (2001). Psychiatric disorders and drug use among human immunodeficiency virus-infected adults in the United States. *Archives of General Psychiatry*, *58*, 721–728.

Burnam, M. A., Bing, E. G., Morton, S. C., Sherbourne, C., Fleishman, J. A., London, A. S., et al. (2001). Use of mental health and substance abuse treatment services among adults with HIV in the United States. *Archives of General Psychiatry*, *58*, 729–736.

Cochran, B. N., & Cauce, A. M. (2006). Characteristics of lesbian, gay, bisexual, and transgender individuals entering substance abuse treatment. *Journal of Substance Abuse Treatment*, *30*, 135–146.

Conigliaro, J., Madenwald, T., Bryant, K., Braithwaite, S., Gordon, A., Fultz, S. L., et al. (2004). The veterans aging cohort study: Observational studies of alcohol use, abuse, and outcomes among human immunodeficiency virus-infected veterans. *Alcoholism-Clinical and Experimental Research*, *28*, 313–321.

Dausey, D. J., & Desai, R. A. (2003). Psychiatric comorbidity and the prevalence of HIV infection in a sample of patients in treatment for substance abuse. *Journal of Nervous and Mental Disease*, *191*, 10–17.

Davis, T., Crouch, M., Long, S., Jackson, R., Bates, P., George, R., et al. (1991). Rapid assessment of literacy levels of adult primary care patients. *Family Medicine Journal*, *23*, 433–435.

Del Boca, F. K., & Darkes, J. (2003). The validity of self-reports of alcohol consumption: State of the science and challenges for research. *Addiction*, *98*, 1–12.

Eaton, W. W., & Kessler, L. G. (1981). Rates of symptoms of depression in a national sample. *American Journal of Epidemiology*, *114*, 528–538.

Ewing, J. A. (1984). Detecting alcoholism—The Cage Questionnaire. *JAMA-Journal of the American Medical Association*, *252*, 1905–1907.

Farley, M., Golding, J. M., Young, G., Mulligan, M., & Minkoff, J. R. (2004). Trauma history and relapse probability among patients seeking substance abuse treatment. *Journal of Substance Abuse Treatment*, *27*, 161–167.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Galvan, F. H., Bing, E. G., Fleishman, J. A., London, A. S., Caetano, R., Burnam, M. A., et al. (2002). The prevalence of alcohol consumption and heavy drinking among people with HIV in the United States: Results from the HIV cost and services utilization study. *Journal of Studies on Alcohol*, *63*, 179–186.

Galvan, F. H., Burnam, M. A., & Bing, E. G. (2003). Co-occurring psychiatric symptoms and drug dependence or heavy drinking among HIV-positive people. *Journal of Psychoactive Drugs*, *35*, 153–160.

Gastfriend, D. R., & McLellan, A. T. (1997). Treatment matching: Theoretical basis and practical implications. *Medical clinics of North America*, *81*, 945–966.

Goldberg, R. W., Seybolt, D. C., & Lehman, A. (2002). Reliable self-report of health service use by individuals with serious mental illness. *Psychiatric Services*, *53*, 879–881.

Goldstein, R. B., Rotheram-Borus, M. J., Johnson, M. O., Weinhardt, L. S., Remien, R. H., Lightfoot, M., et al. (2005). Insurance coverage, usual source of care, and receipt of clinically indicated care for comorbid conditions among adults living with human immunodeficiency virus. *Medical Care*, *43*, 401–410.

Greenfield, S. F., Brooks, A. J., Gordon, S. M., Green, C. A., Kropp, F., McHugh, R. K., et al. (2007). Substance abuse treatment entry, retention, and outcome in women: A review of the literature. *Drug and alcohol dependence*, *86*, 1–21.

Hasin, D. S., & Grant, B. F. (1995). AA and other helpseeking for alcohol problems: Former drinkers in the U.S. general population. *Journal of Substance Abuse*, *7*, 281–292.

Hser, Y. I. (1995). Drug treatment counselor practices and effectiveness. *Evaluation Review*, *19*, 389–408.

Justice, A. C., Homes, W., Gifford, A. L., et al. (2001). Development and validation of a self completed symptom index. *Journal of Clinical Epidemiology*, *54*, 77–90.

Kaskutas, L. A., Bond, J., & Humphreys, K. (2002). Social networks as mediators of the effect of Alcoholics Anonymous. *Addiction*, *97*, 891–900.

Kayo, R., & Rojas, J. (2004). A closer look at how trauma and PTSD can effect substance abuse treatment for women. *Alcoholism-Clinical and Experimental Research*, *28*, 156A.

Kelley, J. L., & Petry, N. M. (2000). HIV risk behaviors in male substance abusers with and without antisocial personality disorder. *Journal of Substance Abuse Treatment*, *19*, 59–66.

Kertesz, S. G., Larson, M. J., Cheng, D. M., Tucker, J. A., Winter, M., Mullins, A., et al. (2006). Need and non-need factors associated with addiction treatment utilization in a cohort of homeless and housed urban poor. *Medical Care*, *44*, 225–233.

Kessler, R., Andrews, G., Mroczeek, D., Ustun, B., & Wittchen, H. (1998). The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI-SF). *International Journal of Methods in Psychiatric Research*, *7*, 171–185.

Kilbourne, A. M., Justice, A. C., Rollman, B. L., et al. (2002). Clinical importance of HIV and depressive symptoms among veterans with HIV infection. *Journal of General Internal Medicine*, *17*, 512–520.

Klinkenberg, W. D., & Sacks, S. (2004). Mental disorders and drug abuse in persons living with HIV/AIDS. *AIDS Care-Psychological and Socio-Medical Aspects of Aids/HIV*, *16*, S22–S42.

Knowlton, A., Arnsten, J., Eldred, L., Wilkinson, J., Gourevitch, M., Shade, S., et al. (2006). Individual, interpersonal, and structural correlates of effective HAART use among urban active injection drug users. *Journal of Acquired Immunodeficiency Diseases*, *41*, 486–492.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied linear regression models*, (4th ed.). New York: McGraw-Hill/Irwin.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data-analysis using generalized linear-models. *Biometrika*, *73*, 13–22.

Loughlin, A., Metsch, L., Gardner, L., Anderson-Mahoney, P., Barrigan, M., & Strathdee, S. (2004). Provider barriers to prescribing HAART to medically-eligible HIV-infected drug users. *AIDS Care-Psychological and Socio-Medical Aspects of AIDS/HIV*, *16*, 485–500.

Lucas, G. M., Gebo, K. A., Chaisson, R. E., & Moore, R. D. (2002). Longitudinal assessment of the effects of drug and alcohol abuse on HIV-1 treatment outcomes in an urban clinic. *AIDS*, *16*, 767–774.

Marlatt, G. A. (1985). Relapse prevention: Theoretical rationale and overview of the model. In G. A. Marlatt, & J. Gordon (Eds.), *Relapse prevention: A self-control strategy for the maintenance of behavior change* (pp. 3–70). NY: Guilford.

Marlatt, G. A. (1996a). Taxonomy of high-risk situations for alcohol relapse: Evolution and development of a cognitive behavior model. *Addiction*, *91*(Suppl 1), S37–S49.

Marlatt, G. A. (1996b). Lest taxonomy become taxidermy: A comment on the relapse replication and extension project. *Addiction*, *91*(Suppl 1), 147–154.

Marlatt, G. A., Barrett, K., & Daley, D. C. (1999). Relapse prevention. In M. Galanter, & H. D. Kleber (Eds.), *Textbook of substance abuse* (2nd ed., pp. 353–365). Washington, DC: American Psychiatric Press.

Masson, C. L., Sorensen, J. L., Phibbs, C. S., & Okin, R. L. (2004). Predictors of medical service utilization among individuals with co-occurring HIV infection and substance abuse disorders. *Aids Care-Psychological and Socio-Medical Aspects of AIDS/HIV*, *16*, 744–755.

Mehta, S. H., Thomas, D. L., Sulkowski, M. S., et al. (2005). A framework for understanding factors that affect access and utilization of treatment for hepatitis C virus infection among HCV-mono-infected and HIV/HCV-co-infected injection drug users. *AIDS*, *19*, S179–S189.

McAuliffe, W. E., & Dunn, R. (2004). Substance abuse treatment needs and access in the USA: Interstate variations. *Addiction*, *99*, 999–1014.

McLellan, A. T., & Weisner, C. (1996). Achieving the public health potential of substance abuse treatment: Implications for patient referral, treatment "matching" and outcome evaluation. In W. Bickel, & R. DeGrandpre (Eds.), *Drug policy and human nature* (pp. 61–80). PA: Williams & Wilkins.

McLellan, A. T., & McKay, J. R. (2003). Components of successful addiction treatment. In A. W. Graham, T. K. Schultz, & M. F. Mayo-Smith, et al. (Eds.), *Principles of addiction medicine* (3rd ed., pp. 429–442). MD: American Society of Addiction Medicine, Inc.

Meyerhoff, D. J. (2001). Effects of alcohol and HIV infection on the central nervous system. *Alcohol Research & Health*, *25*, 288–298.

Moore, R. D., Keruly, J. C., & Chaisson, R. E. (2004). Differences in HIV disease progression by injecting drug use in HIV-infected persons in care. *JAIDS-Journal of Acquired Immune Deficiency Syndromes*, *35*, 46–51.

Najavits, L. M., Crits-Christoph, P., & Dierberger, A. (2000). Clinicians impact on the quality of substance use disorder treatment. *Substance Use & Misuse*, *35*, 2161–2190.

Najavits, L. M., & Weiss, R. D. (1994). Variations in therapist effectiveness in the treatment of patients with substance use disorders: An empirical review. *Addiction*, *89*, 679–688.

Nunes, D., Saitz, R., Libman, H., Cheng, D., Vidaver, J., & Samet, J. H. (2006). Barriers to treatment of hepatitus C in HIV/HCV-coinfected adults with alcohol problems. *Alcoholism: Clinical and Experimental Research*, *30*, 1520–1526.

Orwat, J., Samet, J. H., Tompkins, C. P., Cheng, D. M., Dentato, M. P., & Saitz, R. (2011). Factors associated with attendance in 12-step groups (Alcoholics Anonymous/Narcotics Anonymous) among adults with alcohol problems living with HIV/AIDS. *Drug Alcohol Dependence*, *113*, 165–171.

Orwat, J. (2004). *HIV/AIDS in Massachusetts: A decade of change and challenge.* Boston: Massachusetts Health Policy Forum.

Palepu, A., Raj, A., Horton, N. J., Tibbetts, N., Meli, S., & Samet, J. H. (2005). Substance abuse treatment and risk behaviors among HIV-infected persons with alcohol problems. *Journal of Substance Abuse Treatment*, *28*, 3–9.

Palepu, A., Horton, N. J., Tibbetts, N., Meli, S., & Samet, J. H. (2004). Uptake and adherence to highly active antiretroviral therapy among HIV-infected people with alcohol and other substance use problems: The impact of substance abuse treatment. *Addiction*, *99*, 361–368.

Palepu, A., Tyndall, M., Yip, B., O'Shaughnessy, M. V., Hogg, R. S., & Montaner, J. S. G. (2003). Impaired virologic response to highly active antiretroviral therapy associated with ongoing injection drug use. *JAIDS-Journal of Acquired Immune Deficiency Syndromes*, *32*, 522–526.

Petry, N. M. (1999). Alcohol use in HIV patients: What we don't know may hurt us. *International Journal of STD & AIDS*, *10*, 561–570.

Rabkin, J. G., Ferrando, S. J., Jacobsberg, L. B., & Fishman, B. (1997). Prevalence of Axis I disorders in an AIDS cohort: A cross-sectional, controlled study. *Comprehensive Psychiatry*, *38*, 146–164.

Rabkin, J. G., McElhiney, M. C., & Ferrando, S. J. (2004). Mood and substance use disorders in older adults with HIV/AIDS: Methodological issues and preliminary evidence. *AIDS*, *18*, S43–S48.

Radloff, L. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.

Samet, J. H., Cheng, D. M., Libman, H., Nunes, D. P., Alperen, J. K., & Saitz, R. (2007). Alcohol consumption and HIV disease progression. *Journal of AIDS*, *46*, 194–199.

Samet, J. H., Phillips, S. J., Horton, N. J., Traphagen, E. T., & Freedberg, K. A. (2004). Detecting alcohol problems in HIV-infected patients: Use of the CAGE questionnaire. *AIDS Research and Human Retroviruses*, *20*, 151–155.

Smith, K. L., Horton, N. J., Saitz, R., & Samet, J. H. (2006). The use of mini-mental state examination in recruitment for substance abuse research studies. *Drug and Alcohol Dependence*, *82*, 231–237.

Song, J. (1999). *HIV/AIDS & homelessness.* Nashville, TN: National Health Care for the Homeless Council.

Tsui, J. I., Saitz, R., Cheng, D. M., Nunes, D., Libman, H., Alperen, J. K., et al. (2007). Awareness of hepatitis C diagnosis is associated with less alcohol use among persons co-infected with HIV. *Journal of General Internal Medicine*, *22*, 822–825.

Wadsworth, R., Stampneto, A. M., & Halbrook, B. M. (1995). The role of sexual trauma in the treatment of chemically dependent women: Addressing the relapse issue. *Journal of Counseling and Development*, *73*, 401–406.

Weisner, C., Matzger, H., Tam, T., & Schmidt, L. (2002). Who goes to alcohol and drug treatment? Understanding utilization within the context of insurance. *Journal of Studies on Alcohol*, *63*, 673–682.

Weissman, M. M., Sholomskas, D., Pottenger, M., Prusoff, B. A., & Locke, B. Z. (1977). Assessing depressive symptoms in five psychiatric populations: A validation study. *American Journal of Epidemiology*, *106*, 203–214.

Witbrodt, J., & Kaskutas, L. A. (2005). Does diagnosis matter? Differential effects of 12-step participation and social networks on abstinence. *American Journal of Drug and Alcohol Abuse*, *31*, 685–707.

Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data-analysis for discrete and continuous outcomes. *Biometrics*, *42*, 121–130.

Published in final edited form as:
Drug Alcohol Depend. 2011 January 15; 113(2-3): 165–171. doi:10.1016/j.drugalcdep.2010.07.021.

Factors associated with attendance in 12-step groups (Alcoholics Anonymous/Narcotics Anonymous) among adults with alcohol problems living with HIV/AIDS

John Orwat[1], **Jeffrey H. Samet**[2,3], **Christopher P. Tompkins**[4], **Debbie M. Cheng**[2,5], **Michael P. Dentato**[1], and **Richard Saitz**[2,6]

[1]Loyola University Chicago, School of Social Work, 820 North Michigan Avenue, Chicago, IL, 60611

[2]Boston University School of Medicine Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center, Crosstown, 801 Massachusetts Ave, 2nd Floor, Boston MA 02118

[3]Boston University School of Public Health, Department of Social and Behavioral Sciences, 715 Albany Street, Talbot Building, Boston, MA 02118

[4]Brandeis University, Heller School for Social Policy & Management, 415 South Street, MS 035, Waltham, MA 02454

[5]Boston University, Department of Biostatistics, Graduate School of Arts & Sciences, 705 Commonwealth Avenue, Boston, MA 02215

[6]Boston University, Department of Epidemiology and Youth Alcohol Prevention Center, 801 Massachusetts Ave., Floor 2, Boston, MA 02118.

Abstract

Background—Despite the value of 12-step meetings, few studies have examined factors associated with attendance among those living with HIV/AIDS, such as the impact of HIV disease severity and demographics.

Objective—This study examines predisposing characteristics, enabling resources and need on attendance at Alcoholic Anonymous (AA) and Narcotics Anonymous (NA) meetings among those living with HIV/AIDS and alcohol problems.

Methods—Secondary analysis of prospective data from the HIV-Longitudinal Interrelationships of Viruses and Ethanol study, a cohort of 400 adults living with HIV/AIDS and alcohol problems. Factors associated with AA/NA attendance were identified using the Anderson model for vulnerable populations. Generalized estimating equation logistic regression models were fit to identify factors associated with self-reported AA/NA attendance.

Results—At study entry, subjects were 75% male, 12% met diagnostic criteria for alcohol dependence, 43% had drug dependence and 56% reported attending one or more AA/NA meetings

Corresponding Author: John Orwat, Ph.D. Assistant Professor Loyola University Chicago School of Social Work, Room 1229 820 North Michigan Avenue Chicago, IL 60611 312-915-7494 phone 312-915-7645 fax jorwat@luc.edu.

(past six months). In the adjusted model, female gender negatively associated with attendance, as were social support systems that use alcohol and/or drugs, while presence of HCV antibody, drug dependence diagnosis, and homelessness associated with higher odds of attendance.

Conclusions—Non-substance abuse related barriers to AA/NA group attendance exist for those living with HIV/AIDS, including females and social support systems that use alcohol and/or drugs. Positive associations of homelessness, HCV infection and current drug dependence were identified. These findings provide implications for policy makers and treatment professionals who wish to encourage attendance at 12-step meetings for those living with HIV/AIDS and alcohol or other substance use problems.

Keywords

HIV-infection; alcohol addiction disorder; substance-related disorders; 12 step groups; HIV/AIDS

1. Introduction

Individuals living with the Human Immunodeficiency Virus or Acquired Immune Deficiency Syndrome (HIV/AIDS) are more likely to be diagnosed with alcohol and other drug (AOD) use disorders than those individuals within the general population (Bing et al., 2001; Galvan et al., 2002; Galvan et al., 2003; Rabkin et al., 1997; Rabkin et al., 2004). The impact of AOD use by those living with HIV/AIDS presents the potential for increased challenges upon the health care delivery system (Masson et al., 2004), the heightened likelihood of risky behavior potentially resulting in new HIV infections (Palepu et al., 2003), non-adherence with antiretroviral treatments (ART) (Hendershot et al., 2009; Samet et al., 2004; Conigliaro et al, 2004; Dausey and Desai, 2003; Lucas, et al., 2002), liver complications exacerbated by co-occurring alcohol and other drug use, HIV/AIDS medications, and the hepatitis C virus (HCV) (Lucas et al., 2002; Moore et al., 2004; Palepu et al., 2003; Petry, 1999; Samet et al., 2007). While Alcoholic Anonymous (AA) and Narcotics Anonymous (NA) groups have great potential for the mental and physical health of those living with HIV/AIDS, there are few studies examining factors associated with group meeting attendance.

AA/NA meetings are the most frequently utilized recovery resource for people with AOD related problems (Narrow et al., 1993; Weisner et al., 2002; Weisner and Schmidt, 2001) and more people with drinking disorders attend AA meetings than any other recovery resource (Weisner et al., 1995; Humphreys et al., 1998). As AA/NA groups are free and widely available, they are also associated with positive outcomes, making an independent contribution to the reduction of AOD use, higher levels of abstinence and reduced levels of drinking, lower levels of incarceration, greater psychological adjustment, and lower overall treatment cost (Watson et al., 1997; Timko et al., 2000; Fiorentine and Hillhouse, 2000; Humphreys and Moos, 2001; Kaskutas et al. 2002; Kissin et al., 2003; McCrady and Share, 2003). AA/NA groups may occur in conjunction with formal systems of treatment or mandated requirements by the court system (Speiglman, 1994; Wild et al., 2002). For many, attendance is life long (Fiorentine and Hillhouse, 2000). Over the past decade, research on the process of change and outcomes associated with mutual help groups such as AA/NA supports their significance as a resource for recovery, providing guidelines for living, increasing social support networks and linkage to ongoing care (McCrady and Share, 2003).

Meta-analyses or reviews of multiple studies (Emrick et al., 1993; Tonigan et al., 1996; Bogenschultz, 2008) have reported on sociodemographic factors (e.g., spirituality, race, gender) associated with AA/NA participation. Attendance is more common among racial and ethnic minority groups, those with lower incomes, and those with unstable employment,

suggesting that economic barriers may preclude formal treatment in favor of AA/NA attendance (Humphreys et al., 1998; Kaskutas et al., 1999). Prior formal substance abuse treatment is positively associated with AA/NA attendance, likely due to exposure to these groups (Humphreys et al., 1998; McCrady and Share, 2003).

Few studies of the factors associated with attendance among those living with HIV/AIDS have been published. An early paper addressed the impact of self help on HIV risk reduction, finding that in a sample of injection drug users, those who attended self help were almost twice as likely to reduce or eliminate the risk of HIV than those who did not attend (Sibthorpe, Fleming, and Gould, 1994). This study demonstrates that those attending self-help are amenable to HIV risk reduction interventions and the important role self-help can play in reducing risk. Among those living with HIV/AIDS, Burnam, et al. (2001) examined the factors associated with AA/NA attendance in the HIV Cost and Services Utilization Study (HCSUS), a national probability survey of 2,864 adults living with HIV/AIDS and receiving medical care in the United States in 1996. Controlling for need, African Americans were more likely than Whites to attend AA/NA group meetings, as were those in larger metropolitan areas and those in the Northeast, although sex, age, employment, income, and insurance were not associated with attendance. HIV clinical stage and symptom burden were not associated with attendance but lower CD4 cell count was positively associated with attendance.

This study advances previous research by examining factors associated with attendance at AA/NA group meetings among people living with HIV/AIDS in the era of antiretroviral treatments. In addition, the data set provides the opportunity to explore the impact of other factors, such as social supports, co-occurring depression, and, of particular interest, hepatitis C in a cohort with current or past alcohol problems.

2. Methods

We analyzed data collected for the HIV-Longitudinal Interrelationships of Viruses and Ethanol (HIV-LIVE) study, a prospective cohort study of 400 adults living with HIV/AIDS and alcohol problems. Eligibility criteria for this study included documentation of HIV infection, current or past alcohol problems defined as two or more positive responses to the CAGE (Cut down, Annoyed, Guilty, and Eye opener) alcohol screening questionnaire (Ewing, 1984; Samet et al., 2004) or an alcohol use disorder as assessed by a physician investigator, the ability to speak English or Spanish, a score of 21 or greater on the Mini Mental State Examination (Folstein et al., 1975; Smith et al., 2006) and plans to reside in the area for at least one year. The current analysis included subjects with two or more study interviews. All subjects provided written informed consent and were 18 years of age or older. A certificate of confidentiality was provided by NIAAA as an additional protection of subject confidentiality.

Recruitment of subjects for this study (n = 400) came from two urban hospital-based outpatient practices (30%) in Boston, MA as well as by referral of Boston Medical Center's (BMC) practicing physicians, flyers posted in the community, medical clinics, and addiction treatment facilities (32%). Approximately 38% of participants were recruited from previous observational studies conducted by the investigators (Samet et al., 2007). After obtaining informed consent, subjects enrolled during the period August 2001 until July 2003. Baseline data were collected, and every 6 months thereafter, up to 42 months ending in March 2006. In addition to demographics, interviewers assessed current and past alcohol/drug use and problems, health care utilization, indicators of HIV disease severity, and AA/NA attendance.

2.1 Dependent Variables

The main dependent variable was at least one contact with AA/NA in the previous six months. Subjects were asked the following: “During the past 6 months, did you attend AA or NA meetings?” Secondly, subjects were asked: “How often did you generally attend?” This was followed by prompts for daily, several times a week, weekly, every two weeks, monthly, less than once a month, or other. The second dependent variable was at least weekly attendance over the previous six months, which may approximate a higher level of engagement in self help. In the absence of clear and consistent guidelines for people with substance use disorders generally, we make this approximation.

2.2 Independent variables

were selected using the Andersen Behavioral Model for Vulnerable Populations (Andersen, 1995; Weisner et al., 2002). This model is commonly used in studies to understand utilization of health care by examining three individual level domains: those factors that predispose people to utilize care (e.g., demographics), factors that enable use (e.g., ability to pay), and factors associated with need for the specific service (Andersen, 1995). The comprehensive database allowed for the exploration of other factors, which have yet to be reported on in the literature (e.g., HCV antibody status and social supports).

2.3 Predisposing variables

included age (dichotomized at the median, 44.5), sex, race (Black, White or other), marital status (married, single, or partnered but not married), sexual orientation (gay, bisexual, other vs. heterosexual), education level (high school graduate), living alone, born in the USA, sexual or physical abuse or trauma, and literacy level (a score of >60 on the Rapid Estimate of Adult Literacy in Medicine (REALM), indicating literacy at the high school level) (Davis et al., 1991). Sexual or physical abuse or trauma is considered a predisposing variable to remain consistent with previous literature (Burnam et al., 2001) and since it is not an enabling or need/severity factor when examining utilization of 12-step groups as an outcome. HIV/AIDS disease severity was assessed using CD4 cell count (dichotomized at the median), the presence of opportunistic infections or cancers, and an HIV-Symptom Index score, which is a measure of how often and how bothersome a person experiences 20 common HIV symptoms (Justice, et al., 2001; Kilbourne, et al., 2002) and the presence or absence of the hepatitis C Virus (HCV) antibody (testing was done at the first opportunity if results of testing for clinical purposes were not available).

2.4 Enabling resources

included employment status, receipt of government disability income, homelessness (spending one or more nights “on the street, without shelter” in the past 6 months), and whether or not the subject lived with children. Social support for substance abuse was assessed by asking subjects “how many of the people you spend time with” followed by three prompts “currently drink alcohol,” “are currently heavy or problem drinkers,” and “currently use drugs.” To assess social support for sobriety, subjects were asked, “How many of the people that you spend time with support your sobriety or abstinence?” The use of HIV antiretroviral medications was assessed by asking respondents if they were taking any of a list of all available specific medications over the previous six months, using photographic prompts, generic and brand names.

2.5 Need

for AA/NA variables included alcohol dependence in the past six months assessed using the Composite International Diagnostic Interview Short Form (CIDI-SF) (Kessler et al., 1998) and a diagnosis of drug dependence in the previous six months. Depressive symptoms were

assessed using the Center for Epidemiologic Studies Depression scale (CES-D) (Radloff, 1977; Andresen et al., 1994) and a dichotomous variable was constructed based on a cutoff of 16 or greater. This cutoff has been shown to indicate clinically important depressive symptoms in the general population (Eaton and Kessler, 1981; Weissman et al., 1977). Involvement in the criminal justice system was measured by asking respondents if they were in jail during the previous six months.

3. Analysis

To explore the relationship between the various selected factors and AA/NA group attendance, multivariable analysis was conducted using generalized estimating equations (GEE) (Zeger and Liang, 1986; Liang and Zeger, 1986). This approach was chosen to account for correlation in the data due to incorporating repeated measures from the same subject. Models were fit using a logit link, an exchangeable working correlation structure, and the empirical variance estimator was used in all analyses. An advantage of the GEE approach is that the results are robust, i.e. valid estimates can be obtained from the empirical variance estimator even if the correlation structure is misspecified. In addition, the GEE approach is flexible and can accommodate settings where subjects have varying numbers of follow-up assessments. Independent variables were allowed to vary with time as appropriate and, to assess for potential collinearity, we verified that no pairs of independent variables were highly correlated (> 0.40). Preliminary crude analyses were performed that included only a single independent variable. A full multivariable model was then fit including all independent variables in the same model. To ensure that time varying independent variables preceded the episode of attendance, independent variables were “lagged” to predict report of AA/NA meeting attendance at the subsequent interview. If a participant missed a scheduled follow-up interview, then the outcome at the next available interview was used. We therefore also adjusted for duration of time between assessment of independent variables and outcomes. Similar analyses were used to evaluate the secondary outcome, “weekly or more frequent AA/NA attendance.” Analyses were conducted using two-sided tests and a significance level of 0.05. Due to the exploratory nature of the analyses we did not adjust for multiple comparisons. All analyses were conducted using SAS version 9.0 (SAS Institute, Inc., Cary, NC).

4. Results

4.1 Sample Characteristics

The study sample was comprised of 369 subjects who completed at least two research interviews. Of the initial 369 subjects, 91% completed three interviews, 82% (four interviews), 66% (five interviews), 63% (six interviews), 44% (seven interviews), and 6% completed an eighth interview. A majority of the sample was male, average age was 42.7 years, and was racially diverse (42.8% Black, 32.5% White, 24.7% Other) (Table 1). A majority had experienced sexual or physical abuse or trauma (80.4%), and about a third were gay, bisexual, or “some other” sexual orientation. As for HIV status, a minority had opportunitistic conditions, CD4 count averaged 462 cells/mm^3 and over half had a positive hepatitis C antibody test (58%). A majority of the sample studied reported spending time with people who use alcohol and who are supportive of their sobriety (61.7%) while less than half spent time with people who use drugs (47.3%). As for need variables, 12% met the criteria for a current (past 6-month) alcohol use disorder diagnosis, while almost half met the criteria for a current (past 6-month) drug dependence diagnosis (43.1%). At study entry, 58% had attended AA or NA group meetings during the six months prior, 50% of those in the third interview had attended, 48% in the fourth, 51% in the fifth, and 43% in the sixth, 48% in the seventh, and 67% in the eighth interview.

4.2 Predictors of AA/NA Attendance

The unadjusted and full model are displayed in Table 2. In the unadjusted model (Table 2), living alone, being gay, bisexual, or a sexual orientation other than heterosexual, and alcohol or drug use by social supports had a significant negative association with attendance at AA/NA groups, while social support for sobriety, sexual or physical abuse or trauma, and homelessness in the past six months were positively associated with attendance. Need variables associated with greater odds of attendance were having a current drug dependence diagnosis, depressive symptoms, and having been in jail in the past 6 months.

In the adjusted model, being female was associated with lower odds of attendance, as was a social support system that used alcohol and/or drugs. The presence of HCV antibody and homelessness in the previous 6 months were associated with higher odds of attendance at an AA/NA group, as was having a drug dependence diagnosis.

When the dependent variable was defined as AA/NA attendance weekly or more often (Table 3), few differences emerged in the multivariable model except for minor changes in magnitude and significance of odds ratios. Being female and social supports using alcohol or drugs were still associated with lower odds of AA/NA attendance, while the presence of hepatitis C antibody and meeting the criteria for drug dependence were associated with an increased odds. The association between homelessness and AA/NA attendance was attenuated and no longer statistically significant.

5. Discussion

AA/NA group attendance by those living with HIV/AIDS and alcohol problems is impacted by a variety of factors. Factors that facilitate AA/NA group meeting attendance include a drug dependence diagnosis, homelessness, and the presence of HCV antibody. The odds of attendance were lower for subjects who were female or had social supports that use alcohol or other drugs. When AA/NA was defined as weekly or more frequent attendance, results were similar except the effect of homelessness was no longer statistically significant.

This study underscores the unique differences with regard to attendance in AA/NA groups by those living with HIV/AIDS. While Burnam et al., (2001) did not find significance for demographics including certain geographic locations, sex, age, employment, income, and insurance status, this study revealed the relevance of gender, social supports, homelessness, a drug dependence diagnosis and presence of the HCV antibody with regard to a greater likelihood of attendance or non-attendance of AA/NA groups.

Female gender was negatively associated with attendance in AA/NA, and the negative association between gender and AA/NA group attendance found in this study is **not necessarily** congruent with much of the literature. Some population studies show that women are more likely to drop out of AA/NA (Humphreys et al., 1994) and may not participate for reasons related to program structure, challenges related to lack childcare or if they sense AA/NA is punitive and male dominated (Kaskutas, 1994). Other studies demonstrate that while women may drop out of 12 step attendance more than men after 12 months of treatment (Humphreys et al., 1991), they were more likely to attend in the first place (Humphreys et al., 1994). However, in those studies, after the initial dropout period, women were found to attend as regularly as men (Humphreys et al., 1994). Additionally gender and other sociodemographic variables have not been reliably found to be strongly associated with 12-step attendance in prior studies among general groups of substance abusers (Bogenschultz, 2008; Emrick et al., 1993; Tonigan et al., 1996) or among HIV+ substance users (Burnam et al., 2001).

Social supports currently using alcohol or other drugs were negatively associated with AA/NA group attendance. These findings are consistent with general population studies (Hasin and Grant, 1995; Kaskutas et al., 2002; Witbrodt and Kaskutas, 2005; Moos and Moos, 2006; Kaskutas et al., 2009) related to the negative impact of social supports using alcohol and other drugs on the attendance of those attending AA/NA. This study affirms the negative impact of social supports using alcohol and other drugs on those living with HIV/AIDS who likely would attend AA/NA groups.

There is also a strong and positive independent association with co-occurring hepatitis C, which may be the result of a multi-faceted approach to treatment that coordinates efforts and systematically integrates care. Attendance may also be motivated in an effort to abstain from alcohol or other drugs as it may further complicate their existing health conditions such as hepatitis C. The increased likelihood of AA/NA group attendance among those with co-occurring hepatitis C has implications for clinical practice and policy and is important that both medical and addiction treatment staff clearly understand HIV disease progression and a co-occurring diagnosis of hepatitis C.

On the system level, it would be useful to research the specific mechanisms that facilitate greater attendance of AA/NA groups. The strong and positive association of homelessness and attendance at AA/NA groups, may be due to placement of homeless individuals using substances into detoxification programs or residential services with programmatic requirements as a condition of entry or continuation in the shelters (Kertesz, et al., 2006). Such implications may include exploration of expansion of social networks for the homeless or holding AA/NA groups at convenient locations such as a shelter. Homeless and other low-income persons may be more likely than middle or high income persons to participate in cost-free services for several reasons (e.g., lack of access to other services).

There was also a positive association of those with a drug dependence diagnosis and increased attendance of AA/NA groups while findings were not as significant for those with an alcohol diagnosis or depressive symptoms. This finding, while consistent with other studies (Kaskutas et al., 2009) is curious, and may be related to the fact in addition to seeking specific care such as pharmacotherapy or detoxification, that those with a drug dependence diagnosis are more likely to seek attendance and maintain additional support during and after treatment through AA/NA groups (Timko et al., 2006).

This study offers insight with regard to the factors that may lead to AA/NA group attendance among those living with HIV/AIDS, particularly complementing the literature for those with long-term alcohol and substance use disorders. An additional strength of this study is that the data are quite detailed, which allows for the further examination of the negative impact of social supports that use alcohol or drugs, the role of gender, homelessness, a drug dependence diagnosis, and co-occurring hepatitis C status on AA/NA group attendance. However, limitations of the study are also worth noting. First, the analysis could be strengthened by considering contextual and ecological factors that play a role in participation in AA/NA for those living with HIV/AID, such as the availability of meetings as well as the degree to which the values of the individual are congruent with those manifest in the meetings (Mankowski, Humphreys, & Moos, 2001). Such variables, however, were not available in our data so these analyses were not performed. Second, generalizability of results is limited since data used for this study **were** collected on people living with HIV/AIDS and alcohol problems in the Boston area. The findings are likely applicable to people with HIV/AIDS in similar urban locations. In addition, the observed associations while informative may not be causal.

Reliance on self-reported data is another limitation of this study, which may lead to measurement bias, in that subjects may not accurately report specific behaviors or AA/NA group attendance. Studies do show that self-report is consistent when respondents are asked about service utilization, but may be less reliable when assessing quantity of services (Goldberg et al., 2002). Bias in self-report may have been attenuated in this study by use of validated interview questions as well as the minimization of contextual factors that may influence the possibility of biased self report (Del Boca and Darkes, 2003), along with trained research assistants who emphasized confidentiality and took a systematic approach to interviewing, and a certificate of confidentiality provided by NIAAA as an additional protection of subject privacy.

Despite these limitations, this study offers valuable insight into the factors leading to attendance of AA/NA groups among a cohort of people living with HIV/AIDS. Results have relevance for administrators, clinicians and direct service providers working within the realms of HIV/AIDS, alcohol and other drug treatment, and AA/NA groups. Few data are available about this hard to reach population and AA/NA group attendance. This study has examined these questions using a rich data set that includes detailed predisposing, enabling, and need variables and has made use of clinical data.

This study examined the individual level factors associated with AA/NA group attendance among a cohort of adults living with HIV/AIDS. Of particular interest are variables that are barriers to AA/NA group attendance, such as gender and social supports' use alcohol and drugs. A strong positive association of other variables, such as a drug dependence diagnosis, co-occurring hepatitis C and homelessness indicates the possibility that current policies and practices by medical professionals or amongst service providers are encouraging AA/NA group attendance. Alternatively, the medical challenge of a hepatitis C diagnosis may be a motivator to seek care in the form of AA/NA for one's addiction. Understanding the factors that facilitate or impede AA/NA group attendance may improve the development of strategies to increase attendance, service delivery, referral and care, which is of critical importance for most vulnerable populations.

More research concerning barriers for women and AA/NA attendance is worthwhile since AA/NA is valuable for women in maintaining their sobriety (Beckman, 1993) for many of the same reasons men find AA/NA useful, such as fellowship, group support, and guidance (Kaskutas, 1994). Strategies to increase attendance of women in self help programs may include: increased referrals by medical practitioners to specialty treatment, self help programmatic support to meet the needs of subgroups of women who may have child care challenges or would benefit from meetings for women only, encouragement of female mentorship and/or sponsorship early in the self help process, while paying particular attention to other forms of self-help in the community that may assist in maintaining sobriety. This important issue warrants further exploration, especially in light of the increasing prevalence of HIV/AIDS among ethnic and racial minorities, and most notably women (Orwat, 2004).

There are several potential solutions to increase access and attendance among the specific populations identified with AA/NA groups. Much of this will certainly depend upon the role of medical, mental health and addiction professionals' successful ability to assess needs and link individuals into the appropriate group. However, additional factors must be considered with regard to the elimination of barriers, such as how to work with individuals that have social supports that are using alcohol or other drugs and creating barriers to access and attendance of supportive services such as AA/NA groups.

References

Andersen RM. Revisiting the behavioral-model and access to medical care - Does it matter. J Health Soc Behav. 1995; 36(1):1–10. [PubMed: 7738325]

Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults-Evaluation of a short-form of the CES-D. Am J Prev Med. 1994; 10:77–84. [PubMed: 8037935]

Beckman, LJ. Alcoholics Anonymous and gender issues. In: McCrady, BS.; Miller, WR., editors. Research on Alcoholics Anonymous: Opportunities and Alternatives. Rutgers Center of Alcohol Studies; New Brunswick, NY: 1993.

Bing EG, Burnam A, Longshore D, Fleishman JA, Sherbourne CD, London AS, Turner BJ, Eggan F, Beckman R, Vitiello B, Morton SC, Orlando M, Bozzette SA, Ortiz-Barron L, Shapiro M. Psychiatric disorders and drug use among human immunodeficiency virus-infected adults in the United States. Arch Gen Psychiatry. 2001; 58(8):721–728. [PubMed: 11483137]

Bogenschutz, M. Individual and contextual factors that influence AA affiliation and outcomes. In: Galanter, M.; Kaskutas, L., editors. Recent Developments in Alcoholism: Vol. 18, Research on Alcoholics Anonymous and spirituality in addiction recovery. Springer; Totowa, NJ: 2008. p. 413-428.

Burnam MA, Bing EG, Morton SC, Sherbourne C, Fleishman JA, London AS, Vitiello B, Stein M, Bozzette SA, Shapiro MF. Use of mental health and substance abuse treatment services among adults with HIV in the United States. Arch Gen Psychiatry. 2001; 58(8):729–736. [PubMed: 11483138]

Conigliaro J, Madenwald T, Bryant K, Braithwaite S, Gordon A, Fultz SL, Maisto S, Samet J, Kraemer K, Cook R, Day N, Roach D, Richey S, Justice A. The veterans aging cohort study: Observational studies of alcohol use, abuse, and outcomes among human immunodeficiency virus-infected veterans. Alcohol Clin Exp Res. 2004; 28(2):313–321. [PubMed: 15112939]

Dausey DJ, Desai RA. Psychiatric comorbidity and the prevalence of HIV infection in a sample of patients in treatment for substance abuse. J Nerv Ment Dis. 2003; 191(1):10–17. [PubMed: 12544594]

Davis TC, Crouch MA, Long SW, Jackson RH, Bates P, George RB, Bairnsfather LE. Rapid assessment of literacy levels of adult primary care patients. Fam Med. 1991; 23(6):433–435. [PubMed: 1936717]

Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: State of the science and challenges for research. Addiction. 2003; 98:1–12. [PubMed: 14984237]

Eaton WW, Kessler LG. Rates of Symptoms of Depression in a National Sample. Am J Epidemiol. 1981; 114(4):528–538. [PubMed: 7304583]

Emrick, C.; Tonigan, J.; Montgomery, H.; Little, L. Research on Alcoholics Anonymous: What is currently known. In: McCrady, B.; Miller, W., editors. Research on Alcoholics Anonymous: Opportunities and alternatives. Rutgers Press; New Brunswick, NJ: 1993. p. 41-78.

Ewing JA. Detecting Alcoholism - the Cage Questionnaire. JAMA. 1984; 252(14):1905–1907. [PubMed: 6471323]

Fiorentine R, Hillhouse MP. Drug treatment and 12-step program participation - The additive effects of integrated recovery activities. J Subst Abuse Treat. 2000; 18(1):65–74. [PubMed: 10636609]

Folstein MF, Folstein SE, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975; 12:189–198. [PubMed: 1202204]

Galvan FH, Bing EG, Fleishman JA, London AS, Caetano R, Burnam MA, Longshore D, Morton SC, Orlando M, Shapiro M. The prevalence of alcohol consumption and heavy drinking among people with HIV in the United States: Results from the HIV cost and services utilization study. J Stud Alcohol. 2002; 63(2):179–186. [PubMed: 12033694]

Galvan FH, Burnam MA, Bing EG. Co-occurring psychiatric symptoms and drug dependence or heavy drinking among HIV-positive people. J Psychoactive Drugs. 2003:153–160. [PubMed: 12825758]

Goldberg RW, Seybolt DC, Lehman A. Reliable self-report of health service use by individuals with serious mental illness. Psychiatr Serv. 2002; 53(7):879–881. [PubMed: 12096173]

Hasin DS, Grant BF. AA and other helpseeking for alcohol problems: Former drinkers in the U.S. general population. J Subst Abuse. 1995; 7(3):281–292. [PubMed: 8749788]

Hendershot CS, Stoner SA, Pantalone DW, Simoni JM. Alcohol use and antiretroviral adherence: Review and meta-analysis. J Acquir Immune Defic Syndr. 2009:1–23. (Epub ahead of print).

Humphreys K, Finney JW, Moos RH. Applying a Stress and Coping Framework to Research on Mutual Help Organizations. J Community Psychol. 1994; 22(4):312–327.

Humphreys K, Kaskutas LA, Weisner C. The relationship of pre-treatment Alcoholics Anonymous affiliation with problem severity, social resources and treatment history. Drug Alcohol Depend. 1998; 49(2):123–131. [PubMed: 9543649]

Humphreys K, Mavis BE, Stöffelmayr BE. Factors predicting attendance at self-help groups after substance abuse treatment: Preliminary findings. Journal of Consulting and Clinical Psychology. 1991; 59:591–593. [PubMed: 1918563]

Humphreys K, Mavis BE, Stöffelmayr BE. Are twelve step programs appropriate for disenfranchised groups?: Evidence from a study of posttreatment mutual help group involvement. Prevention in Human Services. 1994; 11:165–180.

Humphreys K, Moos R. Can encouraging substance abuse patients to participate in self-help groups reduce demand for health care? A quasi-experimental study. Alcohol Clin Exp Res. 2001; 25(5): 711–716. [PubMed: 11371720]

Justice AC, Homes W, Gifford AL, Rabeneck L, Zackin R, Sinclair G, Weissman S, Neidig J, Marcus C, Chesney M, Cohn SE, Wu AW, Adult AIDS Clinical Trials Unit Outcomes Committee. Development and validation of a self completed symptom index. J Clin Epidemiol. 2001; 54(1): 77–90.

Kaskutas LA. What do women get out of self-help - Their reasons for attending women for sobriety and Alcoholics-Anonymous. J Subst Abuse Treat. 1994; 11(3):185–195. [PubMed: 8072046]

Kaskutas LA, Keller JW, Witbrodt J. Measuring social model in California: How much has changed? Contemp Drug Probl. 1999; 26:607–631.

Kaskutas LA, Bond J, Humphreys K. Social networks as mediators of the effect of Alcoholics Anonymous. Addiction. 2002; 97(7):891–900. [PubMed: 12133128]

Kaskutas LA, Subbaraman MS, Witbrodt J, Zenmore SE. Effectiveness of making alcoholics anonymous easier: A group format 12-step facilitation approach. J Subst Abuse Treat. 2009; 37:228–239. [PubMed: 19339148]

Kertesz SG, Larson MJ, Cheng DM, Tucker JA, Winter M, Mullins A, Saitz R, Samet JH. Need and non-need factors associated with addiction treatment utilization in a cohort of homeless and housed urban poor. Med Care. 2006; 44(3):225–233. [PubMed: 16501393]

Kessler R, Andrews G, Mroczeek D, Ustun B, Wittchen H. The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI-SF). Int J Methods Psychiatr Res. 1998; 7:171–185.

Kilbourne AM, Justice AC, Rollman BL, McGinnis KA, Rabeneck L, Weissman S, Smola S, Schultz R, Whittle J, Rodriques-Barradas M. Clinical importance of HIV and depressive symptoms among veterans with HIV infection. J Gen Intern Med. 2002; 17(7):512–520. [PubMed: 12133141]

Kissin W, McLeod C, McKay J. The longitudinal relationship between self-help group attendance and course of recovery. Eval Program Plann. 2003; 26(3):311–323.

Liang KY, Zeger SL. Longitudinal Data-Analysis Using Generalized Linear-Models. Biometrika. 1986; 73(1):13–22.

Lucas GM, Gebo KA, Chaisson RE, Moore RD. Longitudinal assessment of the effects of drug and alcohol abuse on HIV-1 treatment outcomes in an urban clinic. AIDS. 2002; 16(5):767–774. [PubMed: 11964533]

Mankowski ES, Humphreys K, Roos RH. Individual and contextual predictors of involvement in twelve-step self-help groups after substance abuse treatment. American Journal of Community Psychology. 2001; 29(4):537–563. [PubMed: 11554152]

Masson CL, Sorensen JL, Phibbs CS, Okin RL. Predictors of medical service utilization among individuals with co-occurring HIV infection and substance abuse disorders. AIDS Care. 2004; 16(6):744–755. [PubMed: 15370062]

McGrady, BS.; Share, D. Recent research into twelve step programs. In: Graham, AW.; Shultz, TK.; Mayo-Smith, MF.; Ries, RK.; Wilford, BB., editors. Principles of addiction medicine. American Society of Addiction Medicine; MD: 2003. p. 955-968.

Moore RD, Keruly JC, Chaisson RE. Differences in HIV disease progression by injecting drug use in HIV-infected persons in care. J Acquir Immune Defic Syndr. 2004; 35(1):46–51. [PubMed: 14707791]

Moos RH, Moos BS. Participation in treatment and Alcoholics Anonymous: A 16-year follow-up of initially untreated individuals. J Clin Psychol. 2006; 62(6):735–750. [PubMed: 16538654]

Narrow WE, Regier DA, Rae DS, Manderscheid RW, Locke BZ. Use of services by persons with mental and addictive disorders - Findings from the National Institute of Mental Health epidemiologic catchment area program. Arch Gen Psychiatry. 1993; 50(2):95–107. [PubMed: 8381266]

Nelder JA, Wedderburn WM. Generalized Linear Models. J R Stat Soc Ser A. 1972; 135(3):370.

Orwat, J. HIV/AIDS in Massachusetts: A decade of change and challenge. Massachusetts Health Policy Forum; Boston: 2004.

Palepu A, Tyndall M, Yip B, O'Shaughnessy MV, Hogg RS, Montaner JSG. Impaired virologic response to highly active antiretroviral therapy associated with ongoing injection drug use. J Acquir Immune Defic Syndr. 2003; 32(5):522–526. [PubMed: 12679704]

Petry NM. Alcohol use in HIV patients: What we don't know may hurt us. Int J STD AIDS. 1999; 10(9):561–570. [PubMed: 10492422]

Rabkin JG, Ferrando SJ, Jacobsberg LB, Fishman B. Prevalence of Axis I disorders in an AIDS cohort: A cross-sectional, controlled study. Compr Psychiatry. 1997; 38(3):146–164. [PubMed: 9154370]

Rabkin JG, McElhiney MC, Ferrando SJ. Mood and substance use disorders in older adults with HIV/AIDS: Methodological issues and preliminary evidence. AIDS. 2004; 18:S43–S48. [PubMed: 15075497]

Radloff L. The CES-D Scale: A self-report depression scale for research in the general population. Appl Psychol Meas. 1977; 1:385–401.

Samet JH, Cheng DM, Libman H, Nunes DP, Alperen JK, Saitz R. Alcohol consumption and HIV disease progression. J Acquir Immune Defic Syndr. 2007; 46(2):194–199. [PubMed: 17667330]

Samet JH, Horton NJ, Meli S, Freedberg KA, Palepu A. Alcohol consumption and antiretroviral adherence among HIV-infected persons with alcohol problems. Alcohol Clin Exp Res. 2004; 28(4):572–577. [PubMed: 15100608]

Samet JH, Phillips SJ, Horton NJ, Traphagen ET, Freedberg KA. Detecting alcohol problems in HIV-infected patients: Use of the CAGE questionnaire. AIDS Res Hum Retroviruses. 2004; 20(2):151–155. [PubMed: 15018702]

Sibthorpe B, Fleming D, Gould J. Self-help groups: a key to HIV risk reduction for high risk drug users? Journal of AIDS. 1994; 7(6):592–598.

Smith KL, Horton NJ, Saitz R, Samet JH. The use of mini-mental state examination in recruitment for substance abuse research studies. Drug Alcohol Depend. 2006; 82(3):231–237. [PubMed: 16256278]

Speiglman R. Mandated AA Attendance for Recidivist Drinking Drivers - Ideology, Organization, and California Criminal-Justice Practices. Addiction. 1994; 89(7):859–868. [PubMed: 8081184]

Timko C, Moos RH, Finney JW, Lesar MD. Long-term outcomes of alcohol use disorders: Comparing untreated individuals with those in alcoholics anonymous and formal treatment. J Stud Alcohol. 2000; 61(4):529–540. [PubMed: 10928723]

Timko C, DeBenedetti A, Billow R. Intensive referral to 12-step self-help groups and 6-month substance use disorder outcomes. Addiction. 2006; 101:678–688. [PubMed: 16669901]

Tonigan J, Toscova R, Miller W. Meta-analysis of the Alcoholics Anonymous literature: Sample and study characteristics moderate findings. Journal of Studies on Alcohol. 1996; 57:65–72. [PubMed: 8747503]

Watson CG, Hancock M, Gearhart LP, Mendez CM, Malovrh P, Raden M. A comparative outcome study of frequent, moderate, occasional, and nonattenders of alcoholics anonymous. J Clin Psychol. 1997; 53(3):209–214. [PubMed: 9075048]

Weisner C, Greenfield T, Room R. Trends in the Treatment of Alcohol-Problems in the U.S. General-Population, 1979 through 1990. Am J Public Health. 1995; 85(1):55–60. [PubMed: 7832262]

Weisner C, Matzger H, Tam T, Schmidt L. Who goes to alcohol and drug treatment? Understanding utilization within the context of insurance. J Stud Alcohol. 2002; 63(6):673–682. [PubMed: 12529067]

Weisner, C.; Schmidt, LA. Rethinking access to alcohol treatment. In: Galanter, M., editor. Recent Developments in Alcoholism. Vol. 15. Kluwer Academic/Plenum Publishers; New York: 2001. p. 107-136.

Weissman MM, Sholomskas D, Pottenger M, Prusoff BA, Locke BZ. Assessing Depressive Symptoms in 5 Psychiatric Populations Validation Study. Am J Epidemiol. 1977; 106(3):203–214. [PubMed: 900119]

Wild TC, Roberts AB, Cooper EL. Compulsory substance abuse treatment: An overview of recent findings and issues. Eur Addict Res. 2002; 8(2):84–93. [PubMed: 11979011]

Witbrodt J, Kaskutas LA. Does diagnosis matter? Differential effects of 12-step participation and social networks. Am J Drug Alcohol Abuse. 2005; 31:685–707. [PubMed: 16320441]

Zeger SL, Liang KY. Longitudinal Data-Analysis for Discrete and Continuous Outcomes. Biometrics. 1986; 42(1):121–130. [PubMed: 3719049]

Table 1

Baseline Characteristics of Subjects Recruited for HIV-LIVE Study

	Variable	Study Sample (n=369)
Predisposing Characteristics	Age, years, Mean (SD)	42.7 (7.5)
	Sex (female)	24.9%
	Race (White)	32.5%
	Race (Black)	42.8%
	Race (Other)	24.7%
	Married	6.7%
	High school graduate *(a)*	65.0%
	Lives alone	28.7%
	Born in the USA	87.5%
	Literacy, High school level *(b)*	63.7%
	Gay, Bisexual, Other Sexual Orientation	32.8%
	Any traumatic abuse, ever *(c)*	80.4%
	HIV Disease Severity Variables	
	Any opportunistic conditions *(d)*	16.6%
	HIV Quality of Life Scale, Mean (SD) *(e)*	9.6 (4.9)
	CD4 Count (cells/mm3), Mean (SD)	462.1 (299.4)
	HCV antibody present	58.0%
Enabling Resources	Employment status (unemployed) *(d)*	73.4%
	Social support uses alcohol or drugs	
	Social support drinks alcohol	61.7%
	Social support uses drugs	47.3%
	Social support helps with sobriety	88.8%
	Currently taking anti-HIV medications	63.7%
	Received disability *(d)*	73.4%
	Homeless *(d)*	24.1%
Need/Severity Domain	Alcohol Diagnosis *(d), (f)*	11.8%
	Drug dependence diagnosis *(d)*	43.1%
	Depressive symptoms *(g)*	62.3%
	In jail *(d)*	18.2%

(a) High school graduate (12 or more years of education).

(b) Literacy score > 60 (Davis et al., 1991).

(c) History of sexual or physical trauma or abuse.

(d) In the past 6 months.

(e) Justice et al. 2001, Kilbourne et al., 2002

[f] Alcohol diagnosis assessed using Composite International Diagnostic Interview Short Form (CIDI-SF) (Kessler et al., 1998).

[g] Measure of depressive symptoms where a CESD score ≥16 indicates substantial depressive symptoms.

Table 2

Factors Associated with any AA/NA Meeting Attendance [¥]

		Crude Model	Full Model
	Variable	**Odds Ratio (95% CI)**	**Adjusted Odds Ratio (95% CI)**
Predisposing Characteristics	Age ([a])	1.06 (0.86, 1.32)	1 (0.97, 1.03)
	Sex (female)	0.78 (0.60, 1.01)	**0.55 (0.34, 0.90)** *
	Race (Black) vs. White	1.16 (0.90, 1.48)	1.36 (0.83, 2.21)
	(Other) vs. White	1.17 (0.87, 1.58)	1.01 (0.53, 1.91)
	Marital Status		
	Married vs. Single	0.76 (0.49, 1.19)	0.96 (0.56, 1.66)
	Partnered, not married vs. Single	0.90 (0.72, 1.12)	0.95 (0.73, 1.25)
	High school graduate ([b])	0.88 (0.70, 1.10)	1.09 (0.70, 1.67)
	Lives alone	**0.66 (0.53, 0.83)** ***	0.98 (0.75, 1.28)
	Born in the USA	0.94 (0.67, 1.32)	0.82 (0.39, 1.72)
	Literacy at high school level ([c])	1.01 (0.80, 1.28)	1.28 (0.82, 2.00)
	Gay, Bisexual, Other Sexual Orientation vs. Heterosexual	**0.52 (0.41, 0.66)** ***	0.64 (0.40, 1.01)
	Any traumatic abuse, ever ([d])	**1.40 (1.06, 1.84)** *	1.66 (0.97, 2.84)
	HIV Disease Severity Variables		
	Any opportunistic conditions ([e])	1.04 (0.79, 1.35)	0.76 (0.56, 1.05)
	HIV Qual.of Life Scale ([f])	**1.47 (1.18, 1.84)** ***	1.03 (0.99, 1.07)
	CD4 Count ([g])	0.99 (0.79, 1.23)	1.02 (0.97, 1.08)
	HCV antibody present	**1.88 (1.50, 2.35)** ***	**2.27 (1.44, 3.58)** ***
Enabling Resources	Unemployed ([e])	1.13 (0.88, 1.44)	0.94 (0.68, 1.31)
	Social support uses alcohol or drugs	**0.47 (0.37, 0.61)** ***	**0.59 (0.44, 0.78)** ***
	Social support helps w/sobriety	**2.09 (1.46, 3.00)** ***	1.03 (0.73, 1.45)
	Lives without children	1.25 (0.85, 1.84)	1.06 (0.70, 1.62)
	Currently taking anti-HIV medications	0.90 (0.72, 1.13)	1.03 (0.77, 1.39)
	Received disability ([e])	1.07 (0.81, 1.40)	0.67 (0.44, 1.01)
	Homeless ([e])	**2.70 (2.01, 3.64)** ***	**1.64 (1.16, 2.33)** **
Need/ Severity	Alcohol Diagnosis ([e]), ([h])	1.19 (0.85, 1.66)	0.96 (0.56, 1.66)
	Drug dependence diagnosis	**1.72 (1.35, 2.19)** ***	**1.37 (1.05, 1.79)** *
	Depressive symptoms ([i])	**1.44 (1.15, 1.79)** **	1.04 (0.80, 1.35)
	In jail ([e])	**2.87 (1.98, 4.17)** ***	1.15 (0.73, 1.80)

*p<05

**p<.01

***p<.001

¥Analyses based on 369 subjects and 1,151 observations.

[a]Age is a dichotomous variable at median (44.5 years).

[b]High school graduate is 12 or more years of education)

[c]Literacy score > 60 (Davis et al., 1991).

[d]History of sexual or physical trauma or abuse.

[e]Past 6 months.

[f]Justice et al. 2001, Kilbourne et al., 2002

[g]CD4 count a dichotomous variable at median.

[h]Alcohol diagnosis using Composite International Diagnostic Interview Short Form (CIDI-SF) (Kessler et al., 1998).

[i]Measure of depressive symptoms where a CESD score ≥16 indicates substantial depressive symptoms.

Table 3

Factors Associated with AA/NA Attendance Weekly or More Often [¥]

	Variable	Adjusted Odds Ratio (95% CI)
Predisposing Characteristics	Age *(a)*	1.01 (0.98, 1.04)
	Sex (female)	**0.53 (0.32, 0.89)** *
	Race (Black) vs. White	1.21 (0.74, 1.96)
	(Other) vs. White	0.95 (0.52, 1.75)
	Marital Status (Married vs. Single)	0.60 (0.30, 1.19)
	(Partnered, not married vs. Single)	0.88 (0.67, 1.17)
	High school graduate *(b)*	0.80 (0.52, 1.23)
	Lives alone	0.93 (0.69, 1.25)
	Born in the USA	0.85 (0.42, 1.71)
	Literacy, High school level *(c)*	1.35 (0.86, 2.12)
	Gay, Bisexual, Other Sexual Orientation vs. Heterosexual	0.80 (0.51, 1.27)
	Any traumatic abuse, ever *(d)*	1.26 (0.74, 2.14)
	HIV DISEASE SEVERITY VARIABLES	
	Any opportunistic conditions *(e)*	0.75 (0.55, 1.02)
	HIV Quality of Life Scale *(f)*	1.01 (0.98, 1.04)
	CD4 Count *(g)*	0.99 (0.94, 1.05)
	HCV antibody present	**2.05 (1.31, 3.20)** **
Enabling Resources	Unemployed *(e)*	0.80 (0.58, 1.09)
	Social support uses alcohol or drugs	**0.56 (0.42, 0.74)** ***
	Social support helps with sobriety	1.03 (0.71, 1.50)
	Lives without children	1.22 (0.70, 2.11)
	Currently taking anti-HIV medications	1.03 (0.75, 1.41)
	Received disability *(e)*	0.73 (0.47, 1.13)
	Homeless *(e)*	1.33 (0.94, 1.87)
Need/Severity	Alcohol Diagnosis *(e), (h)*	0.99 (0.57, 1.75)
	Drug dependence diagnosis	**1.63 (1.22, 2.18)** **
	Depressive symptoms *(i)*	0.94 (0.71, 1.24)
	In jail *(e)*	1.48 (0.94, 2.35)

* p<.05

** p<.01

*** p<.001

¥ Analyses based on 369 subjects and 1,151 observations.

(a) Age is a dichotomous variable at median (44.5 years).

(b) High school graduate (12 or more years of education).

(c) Literacy score > 60 (Davis et al., 1991).

(d) History of sexual or physical trauma or abuse.

(e) Past 6 months.

(f) Justice et al. 2001, Kilbourne et al., 2002

(g) CD4 count is a dichotomous variable at median.

(h) Alcohol diagnosis using Composite International Diagnostic Interview Short Form (CIDI-SF) (Kessler et al., 1998).

(i) Measure of depressive symptoms where a CESD score ≥16 indicates substantial depressive symptoms.

The Mechanisms Linking Health Literacy to Behavior and Health Status

Chandra Y. Osborn, PhD, MPH; Michael K. Paasche-Orlow, MD, MA, MPH; Stacy Cooper Bailey, MPH; Michael S. Wolf, PhD, MPH

***Objective*: To examine the mechanisms linking health literacy to physical activity and self-reported health. *Methods*: From 2005-2007, patients (N=330) with hypertension were recruited from safety net clinics. Pathanalytic models tested the pathways linking health literacy to physical activity and self-reported health. *Results*: There were significant paths from health literacy to knowledge (r=0.22, P<0.001), knowledge to self-efficacy (r=0.13, P<0.01), self-efficacy to physical activity (r=0.17, P<0.01), and physical activity to health status (r=0.17, P<0.01). *Conclusions*: Health education interventions should be literacy sensitive and aim to enhance patient health knowledge and self-efficacy to promote self-care behavior and desirable health outcomes.**

***Key words*: health literacy, health behavior, knowledge, self-efficacy, health status**

***Am J Health Behav.* 2011;35(1):118-128**

The association between limited health literacy and poor health has been supported across acute and chronic disease contexts.[1-3] Patients' health literacy skills may be the first step in a chain of factors impacting health outcomes.[4,5] Although an accumulation of literature on the issue has emerged in the last 2 decades, the mechanisms by which health literacy impacts health are still unclear.[5,6]

Chandra Y. Osborn, Assistant Professor of Medicine, Vanderbilt University Medical Center, Department of Medicine, Nashville, TN. Michael K. Paasche-Orlow, Associate Professor of Medicine, Boston University, School of Medicine, Section of General Internal Medicine, Boston, MA. Stacy Cooper Bailey, Program Director, Health Literacy and Learning Program, Northwestern University Feinberg School of Medicine, Institute for Healthcare Studies, Chicago, IL. Michael S. Wolf, Associate Professor of Medicine and Learning Sciences, Northwestern University Feinberg School of Medicine, Institute for Healthcare Studies, Chicago, IL.

Address correspondence to Dr Osborn, Vanderbilt University Medical Center, Center for Health Services Research, 1215 Twenty-First Ave South, Ste 6000, MCE - North Tower, Nashville, TN 37232-8300. E-mail: chandra.osborn@vanderbilt.edu

Conceptual Causal Model Linking Health Literacy to Health

Paasche-Orlow and Wolf propose plausible causal pathways to explain the well-established association between health literacy and health.[5] They describe systematic, interactional, and self-care mechanisms by which limited health literacy is most likely to lead to worse health outcomes.[5] Drawing from the literature, they argue that social factors (eg, income, social support, culture, language), cognitive/physical factors (eg, memory, hearing, vision), and demographic factors (race/ethnicity, education, and age) determine health literacy skills. They then illustrate how limited health literacy might impact health outcomes at 3 distinct points along a continuum of health care, focusing on (1) access and use of

health care,[1,3,7] (2) provider-patient interactions,[8] and (3) patient self-care.[5] At each respective point, Paasche-Orlow and Wolf explore system factors (complexity, acute care orientation, tiered delivery model), provider factors (communication skills, teaching ability, time, patient-centered care), and extrinsic factors (support technologies, mass media, health education, resources) that could modify and/or mediate the relationship between health literacy and health outcomes.[5]

Patient factors are also affected at each point along the continuum of health care.[5] At the point of access and use of health care, limited health literacy could influence a patient's navigation skills, self-efficacy, and/or perceived barriers to care.[9] During provider-patient interactions, limited health literacy could infringe upon a patient's knowledge, beliefs, and/or level of participation in clinical decision making.[8] Limited health literacy could also negatively influence a patient's motivation,[10] problem-solving ability,[11] self-efficacy,[12] and/or knowledge[4,12,13] required to accurately perform self-care behaviors. Paasche-Orlow and Wolf suggest these aforementioned relationships operate in conjunction with system-, provider-, and extrinsic-level factors noted above and that these relationships exist within the causal pathway linking health literacy to health outcomes.[5]

Health Literacy and Hypertension

Across chronic disease contexts, limited health literacy has been consistently related to determinants of self-care behavior (eg, knowledge and self-efficacy noted above),[4,12,13] poor performance of self-care behavior (or the lack thereof),[14] and worse health outcomes.[3] In hypertension, limited health literacy has been associated with poor disease-related knowledge, [4,15,16] poor medication-refill adherence,[17] unreconciled medications (ie, discrepancies in patient self-reported medication use with his or her medical record),[18] and worse blood pressure control.[15] Limited health literacy has also been associated with an increased risk of hypertension, lower levels of physical function and activity, and worse subjective health status.[19]

Hypertension, Self-care, and Health Status

Leading health organizations endorse regular physical activity to prevent and treat hypertension and its complications.[20] Regular physical activity has been shown to lower blood pressure,[21] enhance endothelial vasodilator function,[22] and prevent the development of left ventricular mass among individuals with hypertension.[23] Synthesizing the research noted above and consistent with the Paasche-Orlow and Wolf model, limited health literacy may affect both objective and subjective indicators of health status via compromised physical function and physical activity.[19] Although Paasche-Orlow and Wolf do not explicitly discuss "self-care behaviors" in their framework, they do focus on the "determinants of self-care," which by nature highlights the importance of "self-care behaviors" in the causal chain. Determinants of physical activity in the hypertension literature are also consistent with the Paasche-Orlow and Wolf framework. Specifically, motivation and self-efficacy have explained 44% of the variance in physical activity;[24] and hypertension treatment knowledge has also shown predictive value.[25]

Empirical Validation of the Causal Pathways

The Paasche-Orlow and Wolf framework highlights both promising areas for intervention research and important gaps in our current understanding of the pathways linking health literacy to health outcomes. Their theoretical paper is one step in what needs to be an iterative process of model specification and clarification. Analytic approaches that test, validate, and/or refine the framework are needed to inform professional responses to the widespread problem of limited health literacy.[26] For example, empirical support for the proposed relationships between health literacy, determinants of self-care, self-care behavior, and health outcomes would undoubtedly inform the design and content and improve the efficacy of patient-level interventions to address health literacy.

In this study, we sought to validate one third of the Paasche-Orlow and Wolf framework. We focused solely on patient self-care pathways using data from adults with hypertension who were receiving care at safety net clinics in diverse areas of the United States. These data include variables on patient demographics, health literacy, knowledge, self-efficacy, and

health status. Using a path analytic approach, we were able to test the following hypothesized paths in the Paasche-Orlow and Wolf framework:

1. Patient demographic characteristics (race/ethnicity, education, and age) predict health literacy.
2. Health literacy predicts patient-level determinants of self-care (knowledge and self-efficacy).
3. Patient-level determinants of self-care predict self-care behavior (physical activity).
4. Self-care behavior predicts health status (subjective health).

In this way, the association between health literacy and health is hypothesized to be accounted for by a sequence of intervening variables (knowledge, self-efficacy, and self-care behavior).

METHODS

Setting and Participants

We recruited consecutive patients with diagnosed hypertension at scheduled appointments from 6 primary care safety net clinics in Grand Rapids, Michigan; Chicago, Illinois; and Shreveport, Louisiana. Clinics in Grand Rapids (n=2) and Chicago (n=2) were federally qualified health centers. One clinic in Shreveport was a community health center; and the other, an ambulatory care clinic at a public hospital. The institutional review boards at each location approved the study procedures. Eligible participants were at least 18 years old, had a diagnosis of hypertension in their medical record, and had a clinic appointment between July 2006 and August 2007. Patients were ineligible if they did not speak English or if the clinic nurse determined (by interaction or chart documentation) they were too ill or cognitively impaired to participate. Nurses identified 377 potentially eligible patients scheduled for clinic appointments and referred them to onsite study personnel. Informed consent was obtained on 334 scheduled interviews. Four of these patients did not complete the interview (N=330). A response rate was determined following the American Association for Public Opinion Research (AAPOR) standards, estimating 87.5% of approached eligible patients participated in the study.[27]

Data Collection

Study personnel conducted in-person interviews at the clinics. Information on demographic characteristics, such as education (1= grades 1-8, 2=grades 9-11, 3=high school, 4=>high school), race/ethnicity (0=White, 1=African American), and age (continuous) were collected. Additional measures included health literacy, knowledge, self-efficacy, self-care behavior (physical activity), and health status.

Health Literacy

Health literacy was assessed using the short version of the Test of Functional Health Literacy in Adults (S-TOFHLA). The S-TOFHLA has 2 parts. One part is a reading comprehension assessment that involves passages of text about medical topics from which every fifth to seventh word is omitted. Respondents must select a suitable word to insert in the missing place from 4 multiple-choice options. The second part assesses numeracy by presenting respondents with questions to determine their ability to use and interpret numbers when reading hospital forms and labeled prescription vials. Scores on the S-TOFHLA range from 0 to 100 and can be categorized as follows: inadequate health literacy (0-53 correct answers), marginal health literacy (54-66), and adequate health literacy (67-100). The S-TOFHLA has demonstrated good internal consistency, reliability, and validity as a categorized and continuous measure of health literacy ($\alpha = 0.98$).[3,28-30] We relied on the continuous score to increase the predictive power in our path analytic models.

Knowledge

Hypertension knowledge was assessed by asking participants a series of questions about the characteristics and symptoms of high blood pressure. Fourteen multiple-choice items made up the scale, and a total score was taken from all questions. Participants were asked about a normal blood pressure reading, lifestyle activities that change blood pressure readings, symptoms of high blood pressure, and complications. The scale was developed for use in the Prudential Health Literacy Survey[4] and was derived from a reliable and valid measure of hypertension knowledge ($\alpha = 0.70$).[16]

Self-efficacy

Self-efficacy to manage high blood pressure was assessed by asking patients

how confident they were in (1) doing all the things necessary to manage their blood pressure; (2) judging when changes in their blood pressure mean they should visit a doctor; (3) doing different activities and tasks to mange their blood pressure so as to reduce the need to see a doctor; (4) reducing the emotional distress caused by their blood pressure; and (5) doing things, other than just taking medication, to reduce how much their blood pressure affects their everyday life. Response options were in Likert format ranging from 1 = not at all confident to 10 = totally confident. In the current study, internal consistency reliability for this measure was $\alpha = 0.76$.

Self-care Behavior (Physical Activity)

The frequency of physical activity was measured with a single-item taken from a 2-item scale that measures both the frequency and duration of physical activity over the past 4 weeks.[31] Response options were 1 = never, 2 = only once or twice, 3 = at least once a week, 4 = 3 to 4 times each week, and 5 = every day.

Health Status

Subjective health was assessed with a widely used general self-rated health question that asks respondents to report how they thought their health was in general.[32] Response options were in Likert format ranging from 1 = Poor to 5 = Excellent. This single question has demonstrated strong predictive validity with self-care behaviors,[33] objective indicators of health,[33] and mortality.[34] Relevant to our study is prior work showing that a single item of self-rated health is sensitive to changes in physical activity.[35,36]

Data Analyses

Patients were categorized as having inadequate health literacy (scores 0-53) or having marginal/adequate health literacy (scores 54-100) and compared using independent samples t-tests for continuous variables and chi-square tests for categorical variables. Pathanalytic models specifying the relationships among variables were estimated using AMOS, version 17. Advantages of this procedure include the generality and flexibility of model specification and the ability to assess fit of the hypothesized model to the observed data. Model fit was examined with the comparative fit index (CFI) and root mean error of approximation (RMSEA).[37] CFI values exceeding 0.90 and RMSEA values below 0.08 indicate reasonable model fit.[37] Hypotheses regarding the specific structural relations of the constructs in the model were also evaluated through inspection of the direction and magnitude of the path coefficients.

Two path analytic models were estimated with a correlation matrix generated by 330 cases, a sample size considered to be of adequate power to detect large effects.[38,39] Demographic information (education, race, and age), health literacy, knowledge, self-efficacy, self-care, and subjective health variables were used to estimate one third of the Paasche-Orlow and Wolf framework (ie, patient factors within the self-care domain). Model 1 (the full model) tested whether demographic factors predicted health literacy; whether health literacy predicted determinants of self-care (knowledge and self-efficacy); whether determinants of self-care predicted self-care behavior (physical activity); and whether self-care behavior predicted health status (subjective health). Although we were interested in only these pathways, all potential paths between variables were included to test both those hypothesized to be significant and those hypothesized to be nonsignificant. Model 2 (the trimmed model) omitted all nonsignificant paths from Model 1. Chi-square difference tests were performed between models 1 and 2 to identify the most parsimonious and final model.

RESULTS

Table 1 provides both an overall statistical description of the study population and stratifies the sample by inadequate and marginal/adequate health literacy categories. The mean age of participants was 53.6 years (SD=12.0); 67.9% were female, and 78.5% were African American. Approximately one-third of participants were recruited from each of the study sites: 30.6% from Chicago, 36.1% from Grand Rapids, and 33.3% from Shreveport. The majority of participants (65.8%) were unemployed/retired, and 43.9% did not have any health insurance. Participants reported having hypertension for an average of 11 years (SD=10.0), and 47.6% had chart-confirmed controlled blood pressure.

Thirteen percent of participants reported having left school prior to the ninth

Table 1
Sociodemographic and Clinical Characteristics of Participants Stratified by Health Literacy

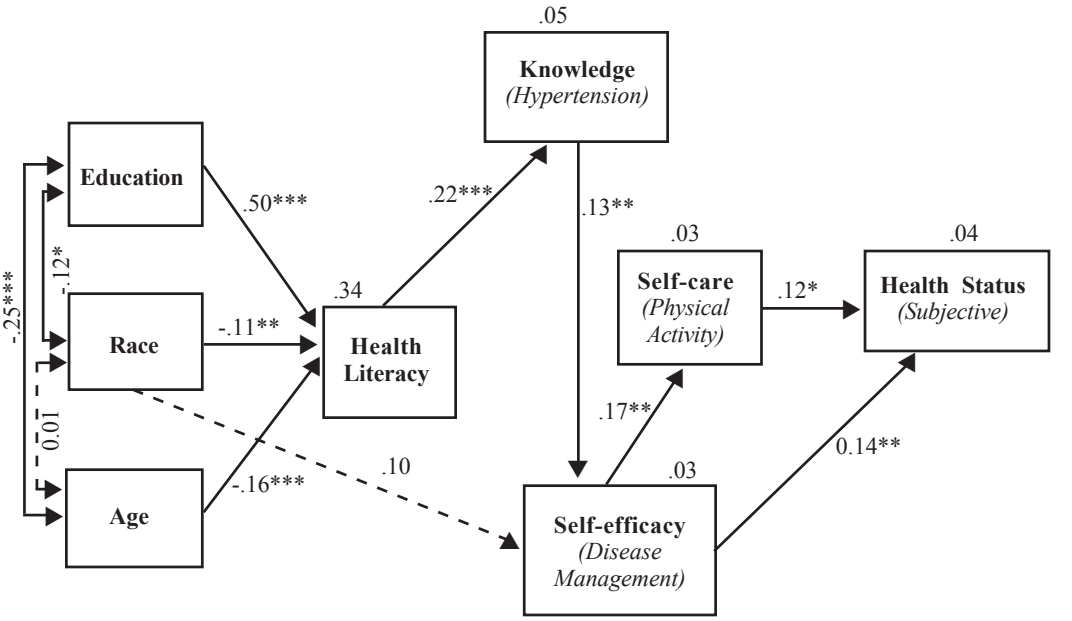
Variable	Total (N=330)	Health Literacy	
		Inadequate (n=100)	Marginal/Adequate (n=230)
Age, Mean (SD)	53.6 (12.0)	58.4 (11.7)**	51.5 (11.6)
Female, %	67.9	63.0	70.6
African American Race, %	78.5	85.9	79.2
Education, %			
Grades 1-8	13.1	29.3***	6.1
Grades 9-11	26.2	41.4***	19.7
High School	31.4	20.2***	36.2
> High School	29.3	9.1***	38
Married, %	30.6	25.0	33.2
Insurance Coverage, %			
Private	10.0	12.0	9.1
Medicare	18.8	19.0	18.7
Medicaid	27.3	29.0	26.5
None/free care	43.9	40.0	45.7
Employment, %			
Full-time	20.9	13.0*	24.3
Part-time	13.3	12.0	13.9
Unemployed/retired	65.8	75.0	61.7
Site, %			
Chicago, IL	30.6	26.0	32.6
Grand Rapids, MI	36.1	38.0	35.2
Shreveport, LA	33.3	36.0	32.2
Years living with hypertension, Mean (SD)	11.0 (10.0)	13.7 (13.1)***	9.9 (8.5)
Knowledge Score, Mean (SD)	11.0 (2.0)	10 (2.5)***	11.3 (2.1)
Self-Efficacy Score, Mean (SD)	37.2 (9.6)	36.3 (9.5)	37.6 (9.6)
Self-care Behavior (Physical Activity) Score, Mean (SD)	3.2 (1.5)	3.2 (1.6)	3.2 (1.5)
Subjective Health Score, Mean (SD)	2.6 (0.9)	2.62 (1.0)	2.6 (0.9)

Note.
Chi-square and student's t-tests for group differences, *P<0.05, **P<0.01 and ***P<0.001.

grade, and another 26.2% had not completed high school. Approximately one third (31.4%) of the sample were high school graduates; fewer years of schooling was significantly associated with inadequate health literacy skills, as well as older age, current unemployment, and greater number of years living with hypertension (Table 1). One third (30.3%) of patients were classified as having inadequate health literacy skills. Means and standard deviations for the knowledge, self-efficacy, self-care behavior (physical activity), and subjective health variables are presented in Table 1.

Model 1 demonstrated excellent data fit, χ^2(3 N=330) = 5.42, P=0.14, CFI=0.99, RMSEA=0.05 (90% CI: 0.00-0.11). There were significant, direct paths from education, race, and age to health literacy (ie, fewer years of education, African American race, and older age were independently associated with lower health literacy scores). Additional significant paths included health literacy to knowledge (ie, higher health literacy scores were associated with more knowledge); self-efficacy to self-care behavior (ie, greater self-efficacy was associated with more physical activity); and self-care be-

Figure 1
Estimated Trimmed Model With Predicted Pathways From Health Literacy to Health Status



Note.
Coefficients are standardized path coefficients. For tests of significance of individual paths, *P<0.05, **P<0.01 and ***P<0.001. Overall model fit, χ^2(16, N=330)=16.75, P=0.40, CFI=1.00, RMSEA=0.01 (90% CI: 0.00-0.05).

havior to health status (ie, more physical activity was associated with more favorable subjective health ratings). In addition, race was directly related to self-efficacy (ie, African American race was associated with greater self-efficacy); and self-efficacy was directly related to health status (ie, greater self-efficacy was associated with more favorable subjective health ratings). All remaining paths were not significant, including the paths from each demographic factor to knowledge, self-efficacy, self-care behavior, or health status; health literacy to self-efficacy, self-care behavior, or health status; and knowledge to self-care behavior or health status.

Model 2 estimated a trimmed version of Model 1, retaining all significant paths and omitting all nonsignificant paths in Model 1. The trimmed model with structural parameters and tests of significance of individual paths appears in Figure 1. Model 2 demonstrated excellent data fit, χ^2(16, N=330)=16.75, P=0.40, CFI=1.00, RMSEA=0.01 (90% CI: 0.00-0.05). All significant paths in Model 1 remained significant in Model 2 except for the path from race to self-efficacy (ie, African American race was no longer directly related to self-efficacy). Demographic factors explained 34% of the variability in the health literacy score, a percentage considered to be a moderate effect for multiple predictor models.[40] Health literacy explained 5% of the variability in knowledge; knowledge explained 3% of the variability in self-efficacy; self-effi-

cacy explained 3% of the variability in self-care behavior (physical activity); and self-efficacy and self-care behavior (physical activity) explained 4% of the variability in health status (subjective health). These are all small effects. A chi-square difference test was performed against Model 1, $\alpha^2(13, N=330)=11.33$, *ns*. The difference was nonsignificant permitting the retention of the trimmed version as the more parsimonious and final model.

DISCUSSION

Conceptual frameworks propose explicit linkages between known determinants of health, with the ultimate goal of locating strategic entry points for intervention. Rooted in prior research, these frameworks require empirical validation prior to their use in developing practical strategies for responding to relevant health problems. We performed a cross-sectional path analysis on data collected from patients with hypertension at safety net clinics in 3 states to evaluate the predicted pathways linking health literacy to health outcomes. Consistent with the Paasche-Orlow and Wolf framework[5] and hypothesis 1, demographic factors (education, race, and age) were directly related to health literacy, were unrelated to all other variables, and explained one third of the variability in health literacy. Consistent with the framework and hypotheses 2-4, health literacy was directly related to knowledge,[4] self-efficacy was directly related to self-care behavior (physical activity),[41,42] and self-care behavior (physical activity) was directly related to health status (subjective health).

Studies in the health literacy, hypertension, and health behavior change literatures support the aforementioned findings. Demographic characteristics, specifically education, race, and age, have been consistently related to health literacy across chronic disease contexts, including hypertension.[43] Prior research has also supported an independent, direct link between health literacy and hypertension knowledge, with limited health literacy being consistently associated with poor knowledge.[4,15,16] Other hypertension studies have supported an independent, direct link between self-efficacy and physical activity,[24,42] as well as an association between physical activity and subjective health.[24] Here, we explored whether knowledge, self-efficacy, and self-care behavior link health literacy to health, which had not been a focus of these latter studies. Although our findings are consistent with prior research, they uniquely extend what is known about the relationship between health literacy and health outcomes. Evidence in support of intervening upon a sequence of modifiable factors in health promotion programs is a significant contribution to both the health literacy and health behavior change literatures and, to our knowledge, has not been done to date.

Inconsistent with the Paasche-Orlow and Wolf framework[5] and hypothesis 2 were nonsignificant paths from health literacy to self-efficacy,[44] self-efficacy to health status, and knowledge to self-care behavior (physical activity). Instead, we found that knowledge was the only significant predictor of self-efficacy, allowing health literacy to be indirectly related to self-efficacy through knowledge. Some studies find health literacy predicts self-efficacy,[14] whereas others have no association.[44] We propose that health literacy may affect self-efficacy through knowledge, but more research is needed to support this. In our study, self-efficacy was an independent, direct predictor of health status. Studies in other chronic disease contexts have shown strong associations between self-efficacy and self-rated health.[45-48] However, less is known about this relationship in the context of hypertension. Lastly, knowledge was unrelated to self-care behavior (physical activity) in our study. This finding is both consistent with behavior change frameworks[49] and prior studies showing knowledge to be poorly correlated with behavior, specifically physical activity.[50,51]

Our findings ruled out several alternative relationships between variables. As illustrated in the Paasche-Orlow and Wolf framework, the impact of demographic factors (education, race, age) on hypertension knowledge, self-efficacy, self-care behavior (physical activity), and health status was entirely explained by health literacy. This is consistent with research showing that health literacy explains demographic differences in self-care behavior and health outcomes.[52,53] In our study, health literacy had no direct relationship with health status. Although this finding is consistent with the Paasche-Orlow and Wolf framework, it is inconsistent with studies in health literacy.[43,54]

One study in particular found a direct link between health literacy and health status that was not mediated by intermediate factors, such as knowledge and behavior.[54] However, because these findings were based on an elderly sample, additional research is needed to rule out cognitive decline as an alternative explanation. Lastly, we found that knowledge had no direct relationship with health status, which has been both theoretically argued[49] and empirically supported in other work.[54]

Several limitations are inherent in this study. First, the data were based on self-reported information, introducing the possibility of deliberate or unconscious misinformation.[55] Furthermore, the measures of self-care behavior (physical activity) and health status were both single-item measures, which have been criticized for introducing bias and not being comprehensive in measurement.[56,57] However, the single-item measures used in this study had been previously used and validated in other research.[31,32] The health status item in particular has demonstrated strong predictive validity with self-care behaviors,[33] objective indicators of health,[33,58] and mortality;[34,59] has been used as a dependent variable in path analytic models;[54,60] and does not violate SEM assumptions for endogenous variables.[38]

Second, although our findings propose causal relationships between variables, the cross-sectional nature of the data precludes causal conclusions and can most appropriately speak to associations between constructs observed at a single point in time. Consequently, we have relied on theory and the research literature to direct our conclusions. Future prospective research is needed to investigate the longitudinal effects of these factors on changes in self-care behaviors and health outcomes.

Third, our findings pertain most directly to populations similar to the participants in this study, and should be confirmed in other populations to see if they generalize. Future work guided by the model proposed in Figure 1 would also provide a more comprehensive conceptualization of the core constructs believed to implicate hypertension self-management.

Finally, although relations between the variables in our study were statistically significant, the magnitude of these relationships was rather modest. Future models should include measures of patient motivation and problem solving; additional self-care behaviors, and objective measures of health status, which may explain more of the variability in the sequence of intervening variables linking health literacy to health outcomes. Duration of and control of high blood pressure are important factors that are likely to impact both health behaviors and health outcomes among patients with hypertension. In an effort to be consistent with the Paasche-Orlow and Wolf framework, duration of disease (eg, hypertension) was not accounted for in our analyses. We also did not include blood pressure control, mainly because it is a very dynamic outcome (ie, patients can be controlled and uncontrolled by turns) – and being in control has a lot to do with activities of the clinicians (eg, providing the proper type and dosage of medication and being available to follow up with patients). We were also unable to classify patients as having prehypertension, stage 1 or stage 2 hypertension.

Despite these limitations, this study is the first to our knowledge to show an indirect pathway from health literacy to health status via widely recognized determinants of self-care (knowledge and self-efficacy) and actual self-care behavior (physical activity). Our findings are just one step in what should be an iterative process of model specification and clarification. Future studies that can both validate our findings and extend them would provide the most useful explanation of the relationship between health literacy and health outcomes and inform professional responses to the problem in many diverse contexts of health care.[61] In the meantime, we suggest, based on our findings, that health literacy-sensitive interventions should aim to enhance disease-specific knowledge that, in turn, will enhance self-efficacy, so that self-efficacy will, in turn, promote the performance of self-care behaviors needed for desirable health outcomes. Future qualitative and quantitative research to determine the most influential techniques to promote knowledge, self-efficacy, and self-care behavior among patients with limited health literacy is needed.

Our findings support the role patient knowledge plays in linking health literacy to health outcomes for patients with hypertension, highlighting the potential impact of patient education efforts as a means of

improving health outcomes. However, intervening to improve knowledge may be challenging. Providers do not have the time. Patients may have other priorities and competing cultural views about hypertension. Effective strategies to promote patient education regarding hypertension for patients with limited health literacy will need to be developed and integrated into care.

Acknowledgments

Funding for this study was provided in part by a research grant awarded to Cherry Street Health Services by the Michigan Department of Community Health. Dr Osborn performed this research under the support of a Diversity Supplement Award (NIDDK P60 DK020593) ■

REFERENCES

1. Baker DW, Parker RM, Williams MV, et al. The relationship of patient reading ability to self-reported health and use of health services. *Am J Public Health.* 1997;87(6):1027-1030.
2. Davis TC, Wolf MS. Health literacy: implications for family medicine. *Fam Med.* Sep 2004;36(8):595-598.
3. Dewalt DA, Berkman ND, Sheridan S, et al. Literacy and health outcomes: a systematic review of the literature. *J Gen Intern Med.* Dec 2004;19(12):1228-1239.
4. Gazmararian JA, Williams MV, Peel J, et al. Health literacy and knowledge of chronic disease. *Patient Educ Couns.* Nov 2003;51(3):267-275.
5. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. *Am J Health Behav.* Sep-Oct 2007;31 (Suppl 1):S19-S26.
6. von Wagner C, Steptoe A, Wolf MS, et al. Health literacy and health actions: a review and a framework from health psychology. *Health Educ Behav.* Aug 2009;36(5):860-877.
7. Beitler JJ, Chen AY, Jacobson K, et al. Health literacy and health care in an inner-city, total laryngectomy population. *Am J Otolaryngol.* Jan-Feb;31(1):29-31.
8. Amalraj S, Starkweather C, Nguyen C, et al. Health literacy, communication, and treatment decision-making in older cancer patients. *Oncology (Williston Park).* Apr 15 2009;23(4):369-375.
9. Baker DW, Parker RM, Williams MV, et al. The health care experience of patients with low literacy. *Arch Fam Med.* Jun 1996;5(6):329-334.
10. Verhoeven L, Snow C, eds. Literacy and Motivation: Reading Engagement in Individuals and Groups. Mahwah, NJ: Lawrence Erlbaum Associates Publishers. (2001) vii, 326 pp.; 2001.
11. Kaminski RA, Good RH, III, eds. Assessing Early Literacy Skills in a Problem-Solving Model: Dynamic Indicators of Basic Early Literacy Skills: Shinn, Mark R; 1998.
12. Torres RY, Marks R. Relationships among health literacy, knowledge about hormone therapy, self-efficacy, and decision-making among postmenopausal health. *J Health Commun.* Jan-Feb 2009;14(1):43-55.
13. Wolf MS, Davis TC, Cross JT, et al. Health literacy and patient knowledge in a Southern US HIV clinic. *Int J STD AIDS.* Nov 2004;15(11):747-752.
14. Wolf MS, Davis TC, Osborn CY, et al. Literacy, self-efficacy, and HIV medication adherence. *Patient Educ Couns.* Feb 2007;65(2):253-260.
15. Pandit AU, Tang JW, Bailey SC, et al. Education, literacy, and health: mediating effects on hypertension knowledge and control. *Patient Educ Couns.* Jun 2009;75(3):381-385.
16. Williams MV, Baker DW, Parker RM, et al. Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. *Arch Intern Med.* Jan 26 1998;158(2):166-172.
17. Gazmararian JA, Kripalani S, Miller MJ, et al. Factors associated with medication refill adherence in cardiovascular-related diseases: a focus on health literacy. *J Gen Intern Med.* Dec 2006;21(12):1215-1221.
18. Persell SD, Osborn CY, Richard R, et al. Limited health literacy is a barrier to medication reconciliation in ambulatory care. *J Gen Intern Med.* Nov 2007;22(11):1523-1526.
19. Kim SH. Health literacy and functional health status in Korean older adults. *J Clin Nurs.* Aug 2009;18(16):2337-2343.
20. Chobanian AV, Bakris GL, Black HR, et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA.* May 21 2003;289(19):2560-2572.
21. Nelson L, Jennings GL, Esler MD, et al. Effect of changing levels of physical activity on blood-pressure and haemodynamics in essential hypertension. *Lancet.* Aug 30 1986;2(8505):473-476.
22. Higashi Y, Sasaki S, Kurisu S, et al. Regular aerobic exercise augments endothelium-dependent vascular relaxation in normotensive as well as hypertensive subjects: role of endothelium-derived nitric oxide. *Circulation.* Sep 14 1999;100(11):1194-1202.
23. Palatini P, Visentin P, Dorigatti F, et al. Regular physical activity prevents development of left ventricular hypertrophy in hypertension. *Eur Heart J.* Jan 2009;30(2):225-232.
24. Lee YS, Laffrey SC. Predictors of physical activity in older adults with borderline hypertension. *Nurs Res.* Mar-Apr 2006;55(2):110-120.
25. Knuth AG, Bielemann RM, Silva SG, et al. [Public knowledge on the role of physical activity in the prevention and treatment of

diabetes and hypertension: a population-based study in southern Brazil]. *Cad Saude Publica.* Mar 2009;25(3):513-520.
26. Olsen J. What characterises a useful concept of causation in epidemiology? *J Epidemiol Community Health.* Feb 2003;57(2):86-88.
27. Johnson TP, Owens L. Survey response rate reporting in the professional literature. In: Proceedings of the 2003 American Statistical Association Conference Section on Survey Research Methods. Alexandria, VA: American Statistical Association; 2004:127–133.
28. Baker DW, Williams MV, Parker RM, et al. Development of a brief test to measure functional health literacy. *Patient Educ Couns.* Sep 1999;38(1):33-42.
29. Parker RM, Baker DW, Williams MV, et al. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. *J Gen Intern Med.* Oct 1995;10(10):537-541.
30. Osborn CY, Weiss BD, Davis TC, et al. Measuring adult literacy in health care: performance of the newest vital sign. *Am J Health Behav.* Sep-Oct 2007;31(Suppl 1):S36-S46.
31. Troped PJ, Saunders RP, Pate RR, et al. Correlates of recreational and transportation physical activity among adults in a New England community. *Prev Med.* Oct 2003;37(4):304-310.
32. Idler EL, Russell LB, Davis D. Survival, functional limitations, and self-rated health in the NHANES I Epidemiologic Follow-up Study, 1992. First National Health and Nutrition Examination Survey. *Am J Epidemiol.* Nov 1 2000;152(9):874-883.
33. Haddock CK, Poston WS, Pyle SA, et al. The validity of self-rated health as a measure of health status among young military personnel: evidence from a cross-sectional survey. *Health Qual Life Outcomes.* 2006;4:57.
34. DeSalvo KB, Bloser N, Reynolds K, et al. Mortality prediction with a single general self-rated health question. A meta-analysis. *J Gen Intern Med.* Mar 2006;21(3):267-275.
35. Rohrer JE, Herman DC, Merry SP, et al. Validity of overall self-rated health as an outcome measure in small samples: a pilot study involving a case series. *J Eval Clin Pract.* Apr 2009;15(2):366-369.
36. Han MA, Kim KS, Park J, et al. Association between levels of physical activity and poor self-rated health in Korean adults: The Third Korea National Health and Nutrition Examination Survey (KNHANES), 2005. *Public Health.* Oct 2009;123(10):665-669.
37. Ridgon E. CFI versus RMSEA: comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling.* 1996;3:369-379.
38. Kline R. Principles and Practice of Structural Equation Modeling. 2nd ed. New York: Guilford 2004.
39. Loehlin JC. Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis. Hillsdale, NJ: Lawrence Erlbaum Associates; 1992.
40. Cohen J. Statistical Power Analysis for Behavioral Sciences (Rev. ed.). New York: Academic Press; 1997.
41. McAuley E, Blissmer B. Self-efficacy determinants and consequences of physical activity. *Exerc Sport Sci Rev.* Apr 2000;28(2):85-88.
42. Cromwell SL, Adams MM. Exercise, self-efficacy, and exercise behavior in hypertensive older African-Americans. *J Natl Black Nurses Assoc.* Jul 2006;17(1):17-21.
43. Hanchate AD, Ash AS, Gazmararian JA, et al. The Demographic Assessment for Health Literacy (DAHL): a new tool for estimating associations between health literacy and outcomes in national surveys. *J Gen Intern Med.* Oct 2008;23(10):1561-1566.
44. DeWalt DA, Boone RS, Pignone MP. Literacy and its relationship with self-efficacy, trust, and participation in medical decision making. *Am J Health Behav.* Sep-Oct 2007;(31 Suppl 1):S27-S35.
45. Sarkar U, Ali S, Whooley MA. Self-efficacy and health status in patients with coronary heart disease: findings from the heart and soul study. *Psychosom Med.* May 2007;69(4):306-312.
46. Salbach NM, Mayo NE, Robichaud-Ekstrand S, et al. Balance self-efficacy and its relevance to physical function and perceived health status after stroke. *Arch Phys Med Rehabil.* Mar 2006;87(3):364-370.
47. Cross MJ, March LM, Lapsley HM, et al. Patient self-efficacy and health locus of control: relationships with health status and arthritis-related expenditure. *Rheumatology (Oxford).* Jan 2006;45(1):92-96.
48. Riazi A, Thompson AJ, Hobart JC. Self-efficacy predicts self-reported health status in multiple sclerosis. *Mult Scler.* Feb 2004;10(1):61-66.
49. Fisher WA, Fisher JD, Harman J. The Information-Motivation-Behavioral Skills Model: a general social psychological approach to understanding and promoting health behavior. In: Suls J, Wallston KA, eds. Social Psychological Foundations of Health and Illness. Malden, MA: Blackwell; 2003:82-106.
50. Hays LM, Clark DO. Correlates of physical activity in a sample of older adults with type 2 diabetes. *Diabetes Care.* May 1999;22(5):706-712.
51. Dishman RK, Sallis JF, Orenstein DR. The determinants of physical activity and exercise. *Public Health Rep.* Mar-Apr 1985;100(2):158-171.
52. Osborn CY, Paasche-Orlow MK, Davis TC, et al. Health literacy: an overlooked factor in understanding HIV health disparities. *Am J Prev Med.* Nov 2007;33(5):374-378.
53. Wolf MS, Knight SJ, Lyons EA, et al. Literacy, race, and PSA level among low-income men newly diagnosed with prostate cancer. *Urology.* Jul 2006;68(1):89-93.

54.Cho YI, Lee SY, Arozullah AM, et al. Effects of health literacy on health status and health service utilization amongst the elderly. *Soc Sci Med.* Apr 2008;66(8):1809-1816.
55.Prince SA, Adamo KB, Hamel ME, et al. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act.* 2008;5:56.
56.Ferrari P, Friedenreich C, Matthews CE. The role of measurement error in estimating levels of physical activity. *Am J Epidemiol.* Oct 1 2007;166(7):832-840.
57.van Doorslaer E, Jones AM. Inequalities in self-reported health: validation of a new approach to measurement. *J Health Econ.* Jan 2003;22(1):61-87.
58.Ried LD, Tueth MJ, Handberg E, et al. Validating a self-report measure of global subjective well-being to predict adverse clinical outcomes. *Qual Life Res.* May 2006;15(4):675-686.
59.Thombs BD, Ziegelstein RC, Stewart DE, et al. Physical health status assessed during hospitalization for acute coronary syndrome predicts mortality 12 months later. *J Psychosom Res.* Dec 2008;65(6):587-593.
60.Smith PM, Frank JW, Mustard CA, et al. Examining the relationships between job control and health status: a path analysis approach. *J Epidemiol Community Health.* Jan 2008;62(1):54-61.Jukkala A, Deupree JP, Graham S. Knowledge of limited health literacy at an academic health center. *J Contin Educ Nurs.* Jul 2009;40(7):298-302;quiz:303-294, 336.

Caring for Patients With Limited Health Literacy

A 76-Year-Old Man With Multiple Medical Problems

Michael Paasche-Orlow, MD, MA, MPH, Discussant

Health literacy is the degree to which individuals have the capacity to obtain, process, and understand health information, skills, and services needed to make informed health decisions and take informed actions. Narratives from Mr J, a 76-year-old man with multiple medical problems and limited health literacy, and his physician exhibit some of the difficulties experienced by patients with limited health literacy. Clinicians can help patients with limited health literacy by removing unneeded complexity in their treatment regimens and in the health care system and by using teach-back methods to assess and improve understanding. Rather than a selective screening approach for limited health literacy, a patient-based universal precaution approach for confirming patient comprehension of critical self-care activities helps ensure that all patients have their health literacy needs identified.

JAMA. 2011;306(10):1122-1129 www.jama.com

DR SHIP: Mr J is a 76-year-old man whose care has been complicated by difficulties understanding his health care and accessing treatment. He lives in the greater Boston area and has Medicare.

Mr J's medical history is significant for hypertension, type 2 diabetes, hyperlipidemia, obesity, and sleep apnea. He has had the same physician for the past 18 years. He was born in South Carolina, completed eighth grade, and then began working. He came to the Boston area about 40 years ago. He has worked a variety of jobs, largely doing manual labor. He stopped working when his vision failed from complications of hypertension, diabetes, cataracts, and a macular hole. When questioned about his ability to read, he invokes limitations due to his visual deficits. Mr J lives alone but has been in a romantic relationship with one woman intermittently for about 5 years. He attends a day program about 3 days a week. He does not smoke or drink alcohol.

Every aspect of Mr J's health and health care has been affected by his limited health literacy. His first visit to his physician of 18 years was after an emergency department visit for hypertensive urgency (blood pressure, 200/100 mm Hg, with visual changes, headache, and weakness) as well as a serum glucose level of 389 mg/dL (21.6 mmol/L).

Mr J's diabetes has been poorly controlled for long stretches of time because of poor adherence to diet and medications. His hemoglobin A_{1c} level has been as high as 14.2% and is currently 8.4%. When his caregivers initially started insulin therapy, they involved his girlfriend (who also has diabetes) in his care. This was successful until their relationship foundered. After another family member was unable to assist, a visiting nurse was brought in to teach him how to self-inject. He has been using a Lantus pen himself since then.

Mr J's challenges have limited the treatment of his severe obstructive sleep apnea, which is associated with short runs of ventricular tachycardia when his oxygen saturation decreases at night. He is not interested in using continuous positive airway pressure (CPAP) and has not kept repeated appointments in the sleep clinic. It is unclear whether he uses nocturnal oxygen and how much.

See also Patient Page.

CME available online at www.jamaarchivescme.com and questions on p 1149.

The conference on which this article is based took place at the Medicine Grand Rounds at Beth Israel Deaconess Medical Center, Boston, Massachusetts, on March 4, 2010.

Author Affiliations: Dr Paasche-Orlow is Associate Professor of Medicine, Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts.
Corresponding Author: Michael Paasche-Orlow, MD, MA, MPH, Boston University Medical Center, 801 Massachusetts Ave, Second Floor, Boston, MA 02118 (mpo@bu.edu).

Clinical Crossroads at Beth Israel Deaconess Medical Center is produced and edited by Risa B. Burns, MD, series editor; Tom Delbanco, MD, Howard Libman, MD, Eileen E. Reynolds, MD, Marc Schermerhorn, MD, Amy N. Ship, MD, and Anjala V. Tess, MD.
Clinical Crossroads Section Editor: Margaret A. Winker, MD, Deputy and Online Editor.

Recently, when Mr J was discharged after a hospitalization for pneumonia, he failed to take the prescribed antibiotic. His nonadherence was identified at his postdischarge visit with his nurse. He stated that he did not fill the prescription because he was told it would cost $98. Investigation by his nurse determined that the prescription would cost less than $2, and he agreed to fill it.

His current medications include amlodipine, 10 mg/d; atenolol, 100 mg/d; gabapentin, 300 mg/d at bedtime; glyburide, 10 mg twice per day; hydrochlorothiazide, 25 mg/d; insulin glargine pen, 10 units/d; lactulose, 2 to 3 tbsp/d as needed for constipation; lisinopril, 40 mg/d; pioglitazone, 45 mg/d; sildenafil, 50 mg before sexual intercourse as needed; simvastatin, 80 mg/d; tamsulosin, 0.4 mg/d; aspirin, 325 mg/d; capsaicin 0.075% cream, 3 to 4 times daily on feet; and terbinafine 1% cream, twice per day. He has no drug allergies.

On examination, Mr J had a blood pressure of 138/64 mm Hg; pulse of 72/min and regular; oxygen saturation of 96% on room air; weight of 215 lb (96.8 kg); and height of 66 in (167.6 cm). He appeared well and in no distress. His lungs were clear bilaterally, with good air movement. His heart examination results were normal, with a regular rate and rhythm, S_1, S_2, and no murmurs. He had no lower extremity edema or skin breakdown.

MR J: HIS VIEW

Well, the only problem I have is in my eyes. I mean, the sugar got my eyes the way it is. So, I mean, nothing can be done about it. I don't forget my medicine. What I mean, it don't do all that much good. I get weak sometimes, and I figure if I get real weak, I'll go and take my medicine. It might help. I mean, I don't know if they're not strong enough or I get the wrong medicine, the wrong kind of medicine. I couldn't say.

I try to take what the doctor prescribes and see whether I'll work with that. And, if that ain't doing too good—a lot of it don't do no good—then I tell him about it. He might change 1 pill.

The doctor give me medicine: "Well, you take this, here, 2 times a day," and so on, so on, so on. Okay, you take it. Tomorrow, you feel the same way. I mean, sometimes you feel worse. Doctors don't explain things very well a lot of times.

When I was a kid, we used to have a lot of colds and the mumps, and all like that. My mother fixed medicine out of different roots out of the ground . . . different types of tea. And that helped a lot. She did the best she could to try to keep us going. And, I mean, here I am.

DR Y: HIS VIEW

Unfortunately, Mr J has limited literacy. And I think, in his case, it's both language and it's health literacy. I believe he's illiterate, actually. We don't actually know when his blood pressure is high or his diabetes is getting out of control. We don't know if he is taking his medications or he's not taking them.

The biggest breakthrough for us has been when his nurse practitioner said, "Oh, why don't we get him prepackaged medications in little blister packs?" which has really been helpful. Because then, he doesn't have to know what he is taking.

He doesn't really feel that it's a priority to prevent heart disease. I don't think, despite numerous discussions, he understands what heart disease is, or understands dialysis, or the many consequences that are associated with chronically elevated blood sugars. He doesn't really get those things.

We didn't see him for a while. He says, "Well, I had to go to the hospital." And I said, "Oh, really? Well, what happened there?" He says, "I don't know. But they let me go, eventually." And, much to my shock, I learned that he had chest pain, required cardiac stenting. And, worse than that, he was supposed to be taking clopidogrel to keep his stents open. He didn't know a thing about this. It was shocking and terrifying.

I don't have a formal way to assess language literacy or general health literacy in my practice. I think that we just need to be conscious that this can be an issue. This is one of the causes of nonadherence to prescribed regimens. And we need to think about it. Should we be assessing our patients for health literacy? How should we be assessing them? And is it worth the time, if it's a time-consuming thing, to assess everyone?

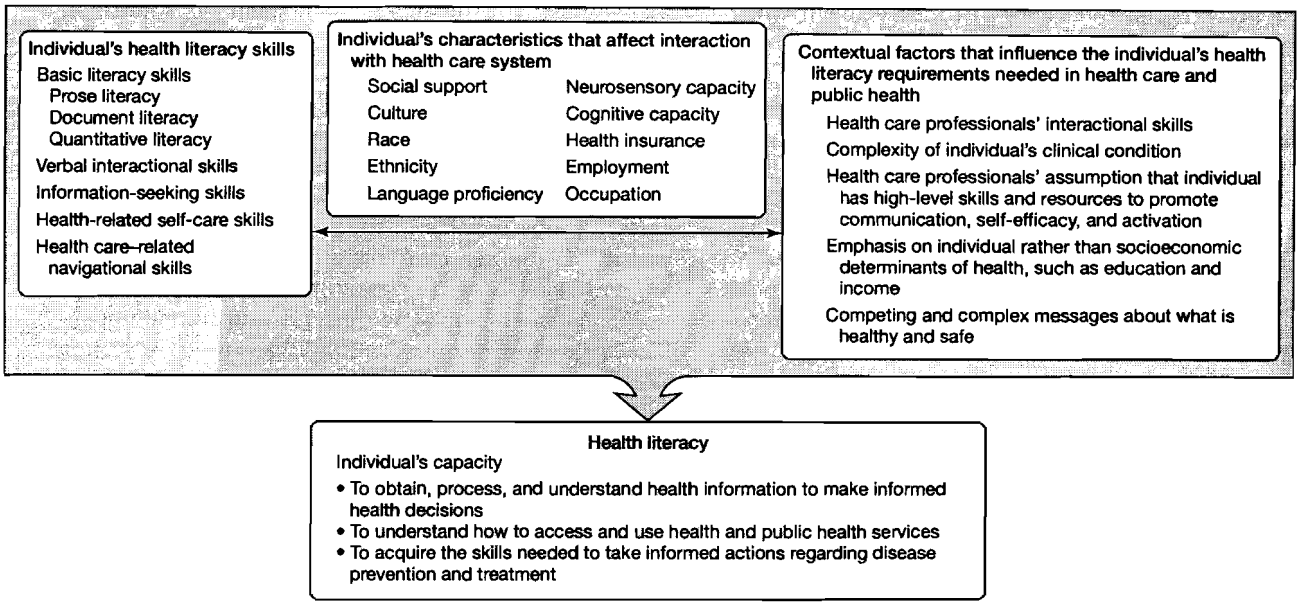
AT THE CROSSROADS: QUESTIONS FOR DR PAASCHE-ORLOW

What are the different conceptual domains that comprise health literacy? What is the "epidemiology" of health literacy issues? To what extent do these issues parallel poor socioeconomic status? What is known about the effects of poor health literacy on patient care or population health? Should clinicians screen for problems with health literacy? If so, how? What do you recommend for Mr J and his caregivers?

Health Literacy

DR PAASCHE-ORLOW: Literacy is inherently a functional concept; ie, it is competence in a set of skills relating to a specific domain of human endeavors. Health literacy is the degree to which individuals have the capacity to obtain, process, and understand health information, skills, and services needed to make informed health decisions and take informed actions. In fact, a broad range of skills is needed to function in relation to one's health. To start, obtaining, processing, and understanding health information often requires the capacity to comprehend written text (prose literacy), forms (document literacy), quantitative information (numeracy), and verbal interactions (interactional skills) (FIGURE). Beyond these fundamental domains of literacy, specific self-care skills are needed according to the tasks that need to be accomplished; eg, inhaler technique or operation of a glucometer. Additional domains of health literacy

Figure. Factors That Contribute to Health Literacy



include information-seeking skills and navigating health systems, although these have been studied less frequently.[1]

Health literacy is itself predicated on a range of linguistic, neurosensory, cognitive, psychiatric, medical, and cultural factors. Although limitations in any of these areas may limit an individual's health literacy and may mediate or moderate the impact of limited health literacy on a person's life, these factors are not themselves domains of health literacy (Figure). For example, patients who have a barrier to comprehension because of low English proficiency deserve language-concordant services and patients who cannot read because of cataracts typically benefit from extraction. It is important to determine patients' specific barriers to health literacy, and interventions should be designed to match a patient's particular issues.

In each domain of health literacy, the types of cognitive challenges faced in health care settings may be quite different from what individuals typically have to manage in their lives. Health care professionals frequently invoke mathematical concepts (eg, risk),[2] complex documents (eg, notices of privacy protection),[3] acronyms, and jargon.[4] Comparisons of different ways to present the rate of benefits or harms from treatment, for example, reveal that even the most successful format for the presentation of rates (as percentages) is misunderstood by one-third of study participants.[5] Even seemingly normal words often have specialized meaning in health settings. For example, health care professionals use the term *diet* to refer to all the calories a person consumes, but most people consider a "diet" to be an organized effort to lose weight. Communication failures are ubiquitous: Is a "negative" biopsy result supposed to be a good thing or a bad thing? Semantic constructions such as "fever spike," "needlestick," and "culture plate" are so routine for health care practitioners that they do not identify these as jargon.

Similarly, health care is replete with experiences that are outside the norm of most people's experience. Medicine's aspiration to promote informed consent and autonomy via shared decision making, although founded on important ethical principles, imparts a significant health literacy burden. For example, the complex issues surrounding prostate cancer screening are difficult to understand. Ultimately, health literacy is a contextually defined phenomenon. Consequently, the Institute of Medicine report on health literacy frames limited health literacy not as a patient problem but as a challenge to health care and public health professionals to communicate with patients more effectively.[6] Specifically, the concept of health literacy should not only incorporate the individual cognitive skills one uses when making health-related decisions but also should take into account the contextual demands placed on the individual by (1) the specific clinical condition; (2) the communication skills of health care professionals;(3) the complex and competing demands of the various health and public health messaging that are encountered; (4) the structure and function of clinical services and public health that assume adequate health literacy and require self-advocacy and vigilance; and (5) the emphasis that society places on individual rather than ecological determinants of health (Figure).

Epidemiology of Health Literacy Issues

The 2003 National Assessment of Adult Literacy (NAAL), the first nationally representative assessment of English health literacy among US adults aged 16 years or older, showed that 14% of the US adult population was found to have below basic health literacy skills and an additional 22% of the US adult population was found to have only basic health

literacy skills.[6] People with below basic health literacy on the NAAL have skills that range from being nonliterate in English to being able to locate easily identifiable information in short, commonplace charts or texts or being able to locate numbers and use them to perform simple operations such as addition when the mathematical information is very concrete and familiar. People with below basic health literacy cannot, for example, use information on the label of an over-the-counter medication to identify substances that may interact to cause an adverse effect. People with basic health literacy skills are able to read and understand information in short, commonplace charts or texts or to locate easily identifiable information and use it to solve simple, 1-step problems when the arithmetic operation is specified or easily inferred. These findings indicate that more than 75 million US adults have limited health literacy skills (ie, below basic or basic on the NAAL).[7]

Surveys of patients' health literacy indicate that the prevalence of limited health literacy is even higher in health settings. In a review of 85 studies from the medical literature including data on 31 129 participants, 46% had limited health literacy.[8] Individuals who are interested in the local prevalence of below basic literacy skills can view state and county estimates at the National Center for Education Statistics Web site.[9]

Effects of Limited Health Literacy

Compared with individuals with adequate health literacy, those with limited health literacy have been shown to have worse health-related knowledge[10] and worse markers of health care processes such as medication adherence,[11] visit adherence,[12] self-care skills,[13,14] intermediate disease markers,[15,16] use of prevention services,[17] delayed diagnoses,[18] and health services utilization.[19] Limited health literacy has also been associated with worse markers of health including health status,[20-22] quality of life,[23,24] hospitalization,[25,26] and mortality.[27-29] For example, in a cohort of 408 English- and Spanish-speaking adults with type 2 diabetes, after adjusting for sociodemographic characteristics, depressive symptoms, social support, treatment regimen, and years with diabetes, individuals with limited health literacy were less likely than those with adequate health literacy to achieve tight glycemic control (hemoglobin $A_{1c} \leq 7.2\%$; adjusted odds ratio [OR], 0.57; 95% CI, 0.32-1.00) and were more likely to have retinopathy (adjusted OR, 2.33; 95% CI, 1.19-4.57).[15] Similarly, in a cohort of 3260 Medicare managed-care enrollees, individuals with limited health literacy had a higher rate of mortality than those with adequate health literacy, with a hazard ratio for all-cause mortality of 1.52 (95% CI, 1.26-1.83) after adjusting for demographics, socioeconomic status, and baseline health.[28] Some reports have presented findings that do not support the relationship between health literacy and health outcomes for topics such as medication adherence and glycemic control.[30,31] The health literacy literature has been reviewed in 2 evidence-based reports presented by the Agency for Healthcare Research and Quality.[32,33]

Socioeconomic Status and Health Literacy

The United States has a significant health literacy gap by educational attainment, income, race, and ethnicity. More than half of African American adults and two-thirds of Hispanic adults have limited health literacy, while less than one-third of white adults have limited health literacy.[7] This gap is parallel to the racial/ethnic gap in general literacy skills, level of educational attainment, and income.[34]

Research has begun to emerge showing how limited health literacy may be an important source of health disparities. Although more research is needed, health literacy has been shown to explain racial disparities in prevention activities,[35] prostate cancer,[36] human immunodeficiency virus (HIV) medication adherence,[37] glycemic control,[38] and end-of-life preferences.[39] For example, in a cohort of 204 persons with HIV infection, health literacy was shown to mediate the observed association between African American race and low medication adherence. In fact, in the final model, the effect of race diminished to nonsignificance and health literacy was the primary predictor of medication nonadherence, such that persons with limited health literacy had a 2.12 (95% CI, 1.93-2.32) higher odds of nonadherence.[37] Such findings suggest that addressing health literacy barriers should help reduce racial/ethnic health disparities.

Screening for Health Literacy

Some have suggested clinical screening for health literacy. Powers et al[40] identified screening tests for reading ability that have been shown to measure literacy with a reasonable degree of accuracy. The validation studies for health literacy screening tools each had their own enrollment criteria to differentiate literacy barriers from visual and cognitive limitations. As such, clinical screening of health literacy should not be performed independently; a positive screening result necessitates additional testing. However, there are reasons to question the underlying premise of clinical screening for health literacy. To my knowledge, the only published trial of screening that assessed clinical outcomes, among patients with diabetes, showed that health literacy screening did not improve outcomes.[41] Studies of patient responses to screening have varied results; patients may have considerable,[42] modest,[43] or low[44] feelings of shame. Regardless, screening should only be performed if there is potential for benefit.[45] In a clinical setting, the most important information to determine is not a health literacy score but whether a patient understands his or her medical conditions, the purpose of the treatment regimen including medications, and how to adhere to the treatment regimen. Other important considerations relate to informed consent for medical procedures. Mr J's literacy screening test result clearly would have been abnormal, but that result would not have addressed his lack of understanding of his illnesses and treatment regimen.

The process of screening for comprehension of the clinical plan has been called "universal precautions for comprehension."[46] Screening a patient for comprehension of the

clinical plan includes identifying any lack of understanding of the plan, simplifying the treatment, and working with the patient until the treatment regimen is understood. For example, in a patient with recalcitrant asthma, mastery of inhaler use is assessed by evaluating the patient's understanding of specific self-care tasks (eg, "Show me which inhaler you should use if you are wheezing. Now show me how you use the inhaler."). This assessment can help direct patient education efforts.

RECOMMENDATIONS FOR MR J AND HIS CAREGIVERS

Mr J has had suboptimal control of his chronic diseases, and his primary care physician has identified limited health literacy as a major cause. In addition to his health literacy, there are additional phenomena that may have impaired Mr J's health care. It is important to consider other types of barriers not only because they may require specific intervention but also because the ensuing evaluation and intervention may be complicated by limited health literacy.

Mr J's history includes his forgetting almost any details of his hospital admission with chest pain that resulted in coronary stent placement and the addition of clopidogrel to his medication list. This episode highlights many additional issues that affect patients' "adherence" to care. First, patient education in transitions of care is notoriously limited. Makaryus and Friedman[47] found that only 42% of patients discharged from the Mayo Clinic could state their diagnosis and even fewer could recall all their medications or common adverse effects. Also, it is common for discharge instructions to lack critical information, to be written in a way that patients do not understand, and not to be sent to primary care clinicians.[48] An alternative possibility for why Mr J did not mention the stent procedure is that he may have been in denial. This is a common phenomenon in coronary artery disease and may have played a role in limiting his self-care activities for medical problems throughout his life.[49]

Important cultural factors may have limited Mr J's health care as well. In many situations, it is difficult to discern between cultural factors and domains of health literacy. For example, understanding what to do with a bottle of prescription medicines requires a number of culturally defined details. The concept of a 30-day supply with refills, the location of this information on the medication label, and how one goes about getting a refill are not standardized.[50] In some respects, persons who do not understand how to interpret medication labels should be regarded as having limited health literacy; however, if such a misunderstanding is due to a lack of familiarity with medical conventions, the issue may need to be regarded as cultural in origin and not due to health literacy. But more fundamental cultural differences can be harder to manage. Mr J appears to take medications for his chronic diseases in a periodic manner in response to symptoms. He is disappointed in the medicines and questions their efficacy when he still feels bad the next day. This pattern of nonadherence may be consistent with not believing in or understanding the concept of chronic asymptomatic disease. The notion of a chronic asymptomatic disease is challenging for many persons because it is not reinforced by personal experience of symptoms as is typical for many other conditions.[51]

To help Mr J, it is vital to understand his cognitive and sensory limitations. Poor vision is actually his chief concern, and this should be tested. Does he have a primary cognitive disability, a dementia process, or pseudodementia? To answer this question, more may be required than performing a Mini-Mental State Examination (MMSE) and depression evaluation. A borderline MMSE score may be confusing because the MMSE score is influenced by education level and limited literacy skills may directly decrease a patient's score (eg, read a sentence, serial 7s).[52] Does he have a psychological (ie, chronic "deniabetes") barrier? Similarly, it would be useful to understand more about his current social milieu. The answers to these questions may alter any other potential interventions.

Mr J should be asked to identify his goals for his medical care, and any gaps that may exist between his goals and his actual self-care activities should be discussed. In this setting, there might be opportunities to examine issues comparing his views with the allopathic model of chronic disease and secondary prevention. Although this approach has not been evaluated in clinical trials, it may lead to an opportunity for his clinicians to compare their goals for Mr J with his own stated goals and potentially to negotiate common ground.

It would be reasonable to ask Mr J to describe his satisfaction with his diabetes care and to compare his comments with Dr Y's degree of satisfaction. This can help clarify differences in their perceptions of how things have been going. While Dr Y is frustrated with nonadherence, Mr J is frustrated with medication adverse effects. It would be good to understand why he forgets to get medications and supplies and why he frequently does not take insulin or other medications. Although evidence has been mixed in trials of patient-centered interviewing to improve diabetes control, such an approach may improve satisfaction and communication regarding adherence.[53]

These approaches may reveal some of the barriers Mr J encounters. Individuals with limited literacy have been shown to be particularly passive in medical encounters.[54] In many clinical scenarios, the default dynamic is that patients need to assert themselves to obtain more information; unfortunately, many patients do not have the self-efficacy required. The concept of universal precautions places a duty on the clinician to affirmatively ascertain patient comprehension.[55]

It is possible that Mr J needs additional training to know what to do. This can be done with a "teach-back" assessment and educational approach,[56] which has been shown to improve asthma self-management[13] and lead to better comprehension of informed consent[57] and to be associated with better metabolic control for patients with diabetes.[58] There are 3 parts to the teach-back. In the first part, the clinician assesses the patient's comprehension (eg, "I want to make sure I explained your medicines well; let's go through each one. I'd like you to tell me how you plan to take each one."). In the second part, the clinician offers feedback that is focused on aspects not understood. In the third part, the clinician reevaluates comprehension ("closes the loop") and provides additional feedback until mastery has been exhibited. Performing the teach-back can help dispel misunderstandings and confirm comprehension but it may also help motivate Mr J in a completely different manner, as this approach exhibits that his clinicians care about him.

It is clear that with support from his former girlfriend, Mr J's medication adherence improved. It is unclear if he has the capacity for independent behavioral change to a life of improved medication adherence. Mr J's care should be made as simple as possible: simplify his medication regimen, expunge all jargon, limit the amount of information discussed per encounter, make a short, action-oriented list of steps Mr J needs to take, review and reinforce the items on his list, and make frequent contact. This is the type of scenario that highlights the potential benefits of a medical home.

It is also appropriate to remember that what Mr J is being asked to do is quite difficult and demanding. Although his medication list is medically reasonable and evidence-based, he is being asked to adhere to a regimen of 16 dose administrations of 12 different medications every day as well as several as-needed prescriptions and to use a CPAP machine. Polypharmacy is an independent risk factor for low adherence.[58] Similarly, adherence to CPAP is notoriously low, with approximately 15% adherence reported among control group participants in a Cochrane review.[59]

Some authors have described warning signs or screening tests that suggest that a patient may have limited health literacy.[60] An alternative approach for clinicians to consider is to examine themselves (and their practice environments) for evidence of unneeded complexity and barriers to effective patient empowerment and education. For example, what are the aspects of your practice that make it hard for patients to ask questions? For a large portion of patients—not just for patients with profound literacy limitations like Mr J—medical practice can be transformed to find ways to elicit questions and concerns and make patient education and empowerment a central activity of patients' health and public health care.[61]

QUESTIONS AND DISCUSSION

QUESTION: Do you think Mr J should find a clinician who speaks his dialect of English?

DR PAASCHE-ORLOW: I do not think he needs a clinician who speaks a different dialect. I would not disrupt the good relationship he has with his physicians but might recommend bringing other individuals into the conversation to see if together Mr J and his girlfriend—and anyone else who could be supportive—might be willing to help him improve his adherence. Although studies about mobilizing family support or peer mentors for diabetes self-care have had mixed outcomes, this still seems like a reasonable approach.[62,63] This would be a big commitment for all involved. Guideline-concordant care for Mr J would likely take several hours a day.[64]

QUESTION: Where would you focus resources? You've talked about both patient factors and basic education as well as physician and system factors. Should clinicians be offering basic health education and literacy as courses for patients or should they focus on patient-physician communication, changing the systems so that they will be accessible and available to limited-literacy patients?

DR PAASCHE-ORLOW: It is hard to choose, and there may be opportunities at all levels. Most of my intervention work to date has been with nonphysician practitioners. It has been much easier for me to train nurses or clinical pharmacists to do the "teach-back" method than to change physicians' behavior. In addition, a lot of documents get thrown at people. Materials need to be markedly simplified and supported by interactive personalized education. I don't think we should regard quality communication as a limited resource. Part of the idea of the medical home is to create a model of care with efficient use of physician extenders to expand patient education and support chronic disease management.

QUESTION: What are your thoughts about the use of multimedia and the Internet with patients?

DR PAASCHE-ORLOW: The first question is "For what purpose?" At this point, there has been an explosion of health information technology activity, but I worry that this is actually likely to increase disparities in the short term.[65] If we can figure out how people with limited health literacy will be able to access such interventions and if we can design easy-to-use interfaces for people with limited health literacy, then maybe we can decrease health disparities down the road. I think it's probably best to use resources to help those who are failing and to focus on the specific issues that each person faces. Multimedia is not always better.[66]

QUESTION: I wonder about closing the loop in the teach-back process you describe. In some ways it can be like an assessment tool when the patient can't close the loop. If you can't close the loop in your session, do you

have a standardized resource or protocol to turn to at that point?

DR PAASCHE-ORLOW: It is quite rare to be unable to close the loop and confirm comprehension. When evaluating comprehension, for instance for use of an inhaler, one sees what the patient's skill level is and provides directed feedback. This typically works. I conducted a study in which we tried to teach to the point of mastery, to close the loop with the inhaler, and almost everyone could be trained.[13] Also, it took about the same number of times around the loop for patients both with limited and with higher health literacy.

Conflict of Interest Disclosures: Dr Paasche-Orlow has completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. He reports receiving funding from multiple grants from the Agency for Healthcare Research and Quality, National Institute on Aging, National Heart, Lung, and Blood Institute, National Cancer Institute, and National Institute on Minority Health and Health Disparities to conduct research in the area of health literacy.
Additional Contributions: We thank the patient and his physician for sharing their stories and for providing permission to publish them.

REFERENCES

1. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. *Am J Health Behav*. 2007;31(suppl 1):S19-S26.
2. Apter AJ, Paasche-Orlow MK, Remillard JT, et al. Numeracy and communication with patients: they are counting on us. *J Gen Intern Med*. 2008;23(12):2117-2124.
3. Paasche-Orlow MK, Jacob DM, Powell JN. Notices of privacy practices: a survey of the Health Insurance Portability and Accountability Act of 1996 documents presented to patients at US hospitals. *Med Care*. 2005;43(6):558-564.
4. Castro CM, Wilson C, Wang F, Schillinger D. Babel babble: physicians' use of unclarified medical jargon with patients. *Am J Health Behav*. 2007;31(suppl 1):S85-S95.
5. Woloshin S, Schwartz LM. Communicating data about the benefits and harms of treatment: a randomized trial. *Ann Intern Med*. 2011;155(2):87-96.
6. Nielsen-Bohlman L, Panzer AM, Kindig DA. *Health Literacy: A Prescription to End Confusion* . Washington, DC: National Academies Press; 2004.
7. Kutner M, Greenberg E, Jin Y, Paulsen C. *The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy (NCES 2006-483)*. Washington, DC: National Center for Education Statistics; 2006.
8. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. *J Gen Intern Med*. 2005;20(2):175-184.
9. National Center for Education Statistics. State and county estimates of low literacy. http://nces.ed.gov/naal/estimates/StateEstimates.aspx. Accessed August 12, 2011.
10. Williams MV, Baker DW, Parker RM, Nurss JR. Relationship of functional health literacy to patients' knowledge of their chronic disease: a study of patients with hypertension and diabetes. *Arch Intern Med*. 1998;158(2):166-172.
11. Kalichman SC, Ramachandran B, Catz S. Adherence to combination antiretroviral therapies in HIV patients of low health literacy. *J Gen Intern Med*. 1999;14(5):267-273.
12. Li BD, Brown WA, Ampil FL, Burton GV, Yu H, McDonald JC. Patient compliance is critical for equivalent clinical outcomes for breast cancer treated by breast-conservation therapy. *Ann Surg*. 2000;231(6):883-889.
13. Paasche-Orlow MK, Riekert KA, Bilderback A, et al. Tailored education may reduce health literacy disparities in asthma self-management. *Am J Respir Crit Care Med*. 2005;172(8):980-986.
14. Sarkar U, Karter AJ, Liu JY, Moffet HH, Adler NE, Schillinger D. Hypoglycemia is more common among type 2 diabetes patients with limited health literacy: the Diabetes Study of Northern California (DISTANCE). *J Gen Intern Med*. 2010;25(9):962-968.
15. Schillinger D, Grumbach K, Piette J, et al. Association of health literacy with diabetes outcomes. *JAMA*. 2002;288(4):475-482.
16. Cavanaugh K, Huizinga MM, Wallston KA, et al. Association of numeracy and diabetes control. *Ann Intern Med*. 2008;148(10):737-746.
17. Scott TL, Gazmararian JA, Williams MV, Baker DW. Health literacy and preventive health care use among Medicare enrollees in a managed care organization. *Med Care*. 2002;40(5):395-404.
18. Bennett CL, Ferreira MR, Davis TC, et al. Relation between literacy, race, and stage of presentation among low-income patients with prostate cancer. *J Clin Oncol*. 1998;16(9):3101-3104.
19. Rothman RL, DeWalt DA, Malone R, et al. Influence of patient literacy on the effectiveness of a primary care-based diabetes disease management program. *JAMA*. 2004;292(14):1711-1716.
20. Wolf MS, Gazmararian JA, Baker DW. Health literacy and functional health status among older adults. *Arch Intern Med*. 2005;165(17):1946-1952.
21. Gazmararian JA, Baker DW, Williams MV, et al. Health literacy among Medicare enrollees in a managed care organization. *JAMA*. 1999;281(6):545-551.
22. Baker DW, Parker RM, Williams MV, Clark WS, Nurss J. The relationship of patient reading ability to self-reported health and use of health services. *Am J Public Health*. 1997;87(6):1027-1030.
23. Tokuda Y, Doba N, Butler JP, Paasche-Orlow MK. Health literacy and physical and psychological wellbeing in Japanese adults. *Patient Educ Couns*. 2009;75(3):411-417.
24. Apter AJ, Wang X, Bogen D, et al. Linking numeracy and asthma-related quality of life. *Patient Educ Couns*. 2009;75(3):386-391.
25. Baker DW, Parker RM, Williams MV, Clark WS. Health literacy and the risk of hospital admission. *J Gen Intern Med*. 1998;13(12):791-798.
26. Baker DW, Gazmararian JA, Williams MV, et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. *Am J Public Health*. 2002;92(8):1278-1283.
27. Sudore RL, Yaffe K, Satterfield S, et al. Limited literacy and mortality in the elderly: the health, aging, and body composition study. *J Gen Intern Med*. 2006;21(8):806-812.
28. Baker DW, Wolf MS, Feinglass J, Thompson JA, Gazmararian JA, Huang J. Health literacy and mortality among elderly persons. *Arch Intern Med*. 2007;167(14):1503-1509.
29. Peterson PN, Shetterly SM, Clarke CL, et al. Health literacy and outcomes among patients with heart failure. *JAMA*. 2011;305(16):1695-1701.
30. Paasche-Orlow MK, Cheng DM, Palepu A, Meli S, Faber V, Samet JH. Health literacy, antiretroviral adherence, and HIV-RNA suppression: a longitudinal perspective. *J Gen Intern Med*. 2006;21(8):835-840.
31. Morris NS, MacLean CD, Littenberg B. Literacy and health outcomes: a cross-sectional study in 1002 adults with diabetes. *BMC Fam Pract*. 2006;7:49.
32. Agency for Healthcare Research and Quality. *Health Literacy Interventions and Outcomes: An Updated Systematic Review*. Rockville, MD: Agency for Healthcare Research and Quality; March 2011. http://www.ahrq.gov/clinic/tp/lituptp.htm. Accessed August 12, 2011.
33. Berkman ND, DeWalt DA, Pignone MP, et al. *Literacy and Health Outcomes*. Rockville, MD: Agency for Healthcare Research and Quality; 2004. Evidence report/technology assessment 87 (AHRQ publication 04-E007-2). http://www.ncbi.nlm.nih.gov/books/NBK37134/. Accessed August 12, 2011.
34. Kutner M, Greenberg E, Jin Y, Boyle B, Hsu Y, Dunleavy E. *Literacy in Everyday Life: Results From the 2003 National Assessment of Adult Literacy*. Washington, DC: National Center for Education Statistics; 2007. NCES publication 2007-480. http://nces.ed.gov/pubs2007/2007480.pdf. Accessed August 12, 2011.
35. Bennett IM, Chen J, Soroui JS, White S. The contribution of health literacy to disparities in self-rated health status and preventive health behaviors in older adults. *Ann Fam Med*. 2009;7(3):204-211.
36. Wolf MS, Knight SJ, Lyons EA, et al. Literacy, race, and PSA level among low-income men newly diagnosed with prostate cancer. *Urology*. 2006;68(1):89-93.
37. Osborn CY, Paasche-Orlow MK, Davis TC, Wolf MS. Health literacy: an overlooked factor in understanding HIV health disparities. *Am J Prev Med*. 2007;33(5):374-378.
38. Osborn CY, Cavanaugh K, Wallston KA, White RO, Rothman RL. Diabetes numeracy: an overlooked factor in understanding racial disparities in glycemic control. *Diabetes Care*. 2009;32(9):1614-1619.
39. Volandes AE, Paasche-Orlow M, Gillick MR, et al. Health literacy not race predicts end-of-life care preferences. *J Palliat Med*. 2008;11(5):754-762.
40. Powers BJ, Trinh JV, Bosworth HB. Can this patient read and understand written health information? *JAMA*. 2010;304(1):76-84.
41. Seligman HK, Wang FF, Palacios JL, et al. Physician notification of their diabetes patients' limited health literacy: a randomized, controlled trial. *J Gen Intern Med*. 2005;20(11):1001-1007.
42. Wolf MS, Williams MV, Parker RM, Parikh NS, Nowlan AW, Baker DW. Patients' shame and attitudes toward discussing the results of literacy screening. *J Health Commun*. 2007;12(8):721-732.
43. Farrell TW, Chandran R, Gramling R. Understanding the role of shame in the clinical assessment of health literacy. *Fam Med*. 2008;40(4):235-236.
44. Ferguson B, Lowman SG, DeWalt DA. Assessing literacy in clinical and community settings: the patient perspective. *J Health Commun*. 2011;16(2):124-134.
45. Paasche-Orlow MK, Wolf MS. Evidence does not support clinical screening of literacy. *J Gen Intern Med*. 2008;23(1):100-102.
46. Paasche-Orlow MK, Schillinger D, Greene SM, Wagner EH. How health care systems can begin to address the challenge of limited literacy. *J Gen Intern Med*. 2006;21(8):884-887.

47. Makaryus AN, Friedman EA. Patients' understanding of their treatment plans and diagnosis at discharge. *Mayo Clin Proc*. 2005;80(8):991-994.
48. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. *JAMA*. 2007;297(8):831-841.
49. Levine J, Warrenburg S, Kerns R, et al. The role of denial in recovery from coronary heart disease. *Psychosom Med*. 1987;49(2):109-117.
50. Standardizing Medication Labels: Confusing Patients Less. Washington, DC: National Academies Press. http://www.iom.edu/Reports/2008/Standardizing-Medication-Labels-Confusing-Patients-Less-Workshop-Summary.aspx. Accessed August 12, 2011.
51. Harvey JN, Lawson VL. The importance of health belief models in determining self-care behaviour in diabetes. *Diabet Med*. 2009;26(1):5-13.
52. Baker DW, Gazmararian JA, Sudano J, Patterson M, Parker RM, Williams MV. Health literacy and performance on the Mini-Mental State Examination. *Aging Ment Health*. 2002;6(1):22-29.
53. Hahn SR, Friedman DS, Quigley HA, et al. Effect of patient-centered communication training on discussion and detection of nonadherence in glaucoma. *Ophthalmology*. 2010;117(7):1339-1347, e6.
54. Katz MG, Jacobson TA, Veledar E, Kripalani S. Patient literacy and question-asking behavior during the medical encounter: a mixed-methods analysis. *J Gen Intern Med*. 2007;22(6):782-786.
55. Volandes AE, Paasche-Orlow MK. Health literacy, health inequality and a just healthcare system. *Am J Bioeth*. 2007;7(11):5-10.
56. Sudore RL, Landefeld CS, Williams BA, Barnes DE, Lindquist K, Schillinger D. Use of a modified informed consent process among vulnerable patients: a descriptive study. *J Gen Intern Med*. 2006;21(8):867-873.
57. Schillinger D, Piette J, Grumbach K, et al. Closing the loop: physician communication with diabetic patients who have low health literacy. *Arch Intern Med*. 2003;163(1):83-90.
58. Bailey CJ, Kodack M. Patient adherence to medication requirements for therapy of type 2 diabetes. *Int J Clin Pract*. 2011;65(3):314-322.
59. Smith I, Nadig V, Lasserson TJ. Educational, supportive and behavioural interventions to improve usage of continuous positive airway pressure machines for adults with obstructive sleep apnoea. *Cochrane Database Syst Rev*. 2009;(2):CD007736.
60. Chew LD, Griffin JM, Partin MR, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. *J Gen Intern Med*. 2008;23(5):561-566.
61. Agency for Healthcare Research and Quality. Health Literacy Universal Precautions Toolkit. Rockville, MD: Agency for Healthcare Research and Quality; April 2010. AHRQ publication 10-0046-EF. http://www.ahrq.gov/qual/literacy/index.html.
62. Rosland AM, Piette JD. Emerging models for mobilizing family support for chronic disease management: a structured review. *Chronic Illn*. 2010;6(1):7-21.
63. Funnell MM. Peer-based behavioural strategies to improve chronic disease self-management and clinical outcomes: evidence, logistics, evaluation considerations and needs for future research. *Fam Pract*. 2010;27(suppl 1):i17-i22.
64. Russell LB, Suh DC, Safford MA. Time requirements for diabetes self-management: too much for many? *J Fam Pract*. 2005;54(1):52-56.
65. Paasche-Orlow MK, Wolf MS. Promoting health literacy research to reduce health disparities. *J Health Commun*. 2010;15(suppl 2):34-41.
66. Wilson EA, Park DC, Curtis LM, et al. Media and memory: the efficacy of video and print materials for promoting patient education about asthma. *Patient Educ Couns*. 2010;80(3):393-398.

Author in the Room Teleconference

Join Dr Paasche-Orlow, the author of this article, on Wednesday, October 19, 2011, from 2 to 3 PM eastern time for "Author in the Room," an interactive teleconference aimed at closing the gap between knowledge—what is published in this article—and action—how this knowledge can be put into practice. This teleconference, facilitated by clinical experts, should help readers answer any questions and consider the implications of the article for their practice.

Author in the Room is brought to you by *JAMA* and the Institute for Healthcare Improvement. To register for Author in the Room, please visit http://www.ihi.org/authorintheroom.You can listen to past conferences or subscribe to the podcast at http://jama.ama-assn.org/authorintheroom/authorindex.dtl.

Journal of Health Communication, 16:328–329, 2011

ISSN: 1081-0730 print/1087-0415 online
DOI: 10.1080/10810730.2011.604393

Routledge
Taylor & Francis Group

Crossword Puzzle: Curl Up with a Good One

MICHAEL K. PAASCHE-ORLOW

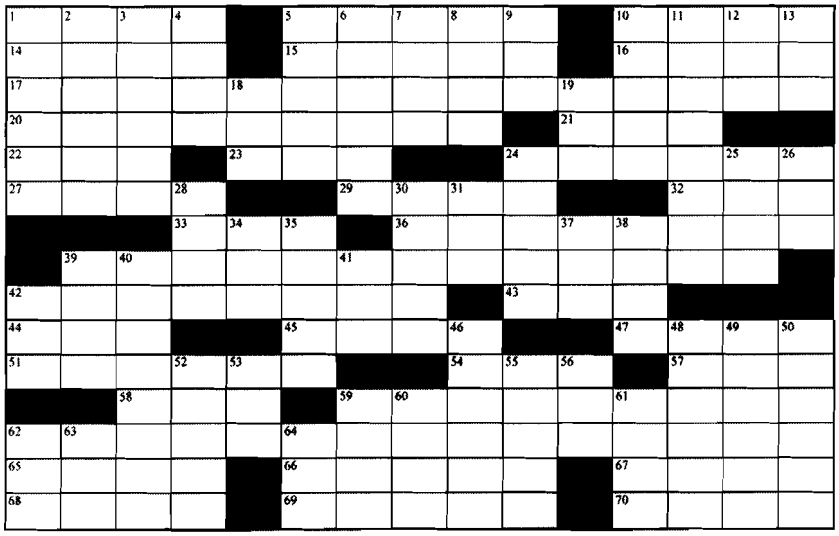
Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

LOUELLA HUNG

Health Research & Educational Trust, American Hospital Association, Chicago, Illinois, USA

First person to submit completed puzzle (see next page for puzzle) will be honored and receive a prize at the Health Literacy Annual Research Conference, 10/17-8, 2011 (http://www.bumc.bu.edu/healthliteracyconference/). Submit to mpo@bu.edu or by fax (617)-414-4676.

Address correspondence to Michael K. Paasche-Orlow, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA. E-mail: mpo@bu.edu



Across:

1. As good as a wink to a blind horse
5. Part of country capital name
10. French ending to establish state
14. Brand of coffee substitute
15. Blue endangered butterfly
16. Pond patina
17. "I find television very educating. Every time somebody turns on the set, ____." Marx quote, Part 1
20. Stop on the Hyannis Ferry
21. Outdoor gear co.
22. Marx quote, Part 3
23. Winnie and Jack's middle name
24. Salad green and mild laxative
27. Marx quote, Part 2
29. Palindromic salutation
32. "Who ___." Song for 24601
33. Theater org.
36. Ante
39. "Wear the old coat and ____", Phelps
42. Paraffin
43. ____ Luis Obispo
44. A vagal nerve joke?
45. Protein source
47. Wyatt
51. Of or relating to embryo, prefix
54. Cause of LBP
57. Type of neck
58. Rosie Cotton's spouse
59. Small town in Idaho
62. Rumsfeld's most mysterious worries
65. Marx quote, Part 4
66. Marx quote, Part 5
67. Gyn procedure
68. ____*jirga*, grand assembly
69. Pardon me in Padua
70. Chelating agnt.

Down:

1. Shaped like a blackberry
2. Olympic host city
3. Washington State town & cherry variant
4. Often preceded by "Just"
5. Eighth letter
6. Imperative to Danno
7. Opposite of new in Nurenberg
8. Dollar in Bangkok
9. Enzyme suffix
10. Ubiquitous compound in fragrances & fabrics
11. Victoria University Cultural Literacy author
12. An angry boy's name
13. Pt. Doc. Sys.
18. Goofball
19. Gold in Guadalajara
24. Cousins of Golden Eyes and Mergansers
25. Valley in Israel
26. Every pot has a ___
28. Aioli
30. Disruption of effective slumber
31. "No matter where you go, there you ___" Buckaroo Banzai
34. Undeliverable mail abbr.
35. Antineoplastic regimen for short
37. Oracle mngr.
38. Tops
39. Turn or tolerate
40. Tries to convince Hutchison how to vote?
41. Flammable finisher
42. Former ISI competition
46. ____ of Titan, Marvel villain
48. Firm believer
49. Curdling agent
50. Poodle Pool?
52. Martial mammal of movies
53. FFS alt.
55. Haley of South Carolina
56. Peptidoglycan abbr.
59. News Ntwk.
60. Vietnamese rice wine
61. Dermatologist's favorite Mexican Sauce?
62. Internet adr.
63. Matrix protagonist
64. What ___ I thinking?

Addictive Behaviors 36 (2011) 539–542

Contents lists available at ScienceDirect

Addictive Behaviors

Short Communication

Using personalized feedback to reduce alcohol use among hazardous drinking college students: The moderating effect of alcohol-related negative consequences

Tibor P. Palfai [a,*], Rebecca Zisserson [a], Richard Saitz [b,c]

[a] Department of Psychology, Boston University, 648 Beacon St., Boston, MA 02215, United States
[b] Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, United States
[c] Youth Alcohol Prevention Center and Department of Epidemiology, Boston University School of Public Health, Boston, MA, United States

ARTICLE INFO

Keywords:
Alcohol
College
Student
Intervention
Web-based
Feedback
Assessment

ABSTRACT

Web-based screening and brief interventions that include personalized feedback about their alcohol use have proven to be particularly promising for reducing hazardous drinking among university students. Despite the increasing use of these approaches, there is still relatively little known about how the content of these interventions may influence outcomes and who may benefit most from these approaches. The current study sought to address these issues by examining how individual differences in alcohol consequences influence outcomes of a laboratory-based computerized intervention.

Methods: One-hundred and nineteen introductory psychology students who either had two episodes of heavy episodic drinking in the past month or scored ≥8 on the AUDIT participated in this randomized controlled trial for course credit. Participants were assigned to 1 of 4 conditions in this 2 Intervention (Alcohol Feedback vs. Control)×2 Assessment (Motivational Assessment vs. No Motivational Assessment) between-subjects design. Quantity of alcohol consumed per week and heavy episodic drinking one month later were the primary dependent variables.

Results: Controlling for corresponding baseline alcohol measures, hierarchical linear regression analyses showed a significant interaction between intervention condition and baseline alcohol-related consequences. For those who reported more alcohol consequences at baseline, the alcohol intervention resulted in significantly less alcohol use and fewer heavy drinking episodes at follow-up, while no difference was observed between intervention conditions for those with few baseline consequences. Assessment did not moderate intervention effects.

Discussion: These findings suggest that a feedback-based computerized intervention that includes normative information about alcohol use and consequences may be more effective for hazardous drinking students who are experiencing higher levels of alcohol-related consequences.

1. Introduction

A large proportion of college students drink in a manner that puts them at risk for alcohol related harm (Hingson, Heeren, Zakocs, Kopstein, & Wechsler, 2002; Wechsler, Lee, Kuo, & Lee, 2000). Such findings have led to increasing efforts to develop effective intervention approaches that may be widely disseminated to a population of drinkers who typically do not seek treatment or identify their drinking behavior as problematic. One of the more promising approaches has been the use of web-based, personalized feedback about alcohol use (Elliott, Carey, & Bolles, 2008). Interventions that have included alcohol feedback (e.g., comparison of participant use to descriptive norms regarding alcohol frequency and quantity) have been shown to be effective at reducing alcohol use among college students (Neighbors, Larimer, & Lewis, 2004; Walters, Vader, & Harris, 2007). In addition to providing corrective information about peer drinking norms in a salient manner, personalized feedback may be used to enhance motivation to change by making individuals aware of the discrepancies between current alcohol use behavior and personal goals, standards, and values (Walters & Neighbors, 2005).

Although there is increasing evidence that students may be responsive to feedback-based interventions for alcohol use, there are a number of unanswered questions regarding how to optimize the efficacy of these approaches. Feedback-based interventions for college students have included different types of information (e.g., drinking norms, costs, and consequences) that have been delivered over varying durations. Despite the wide range of available interventions for college students, there is relatively little known about what information is most effective for promoting change in alcohol use behavior and whether this information may depend on individual differences (Carey, Scott-Sheldon, Carey, & DeMartini, 2007; Elliott et al., 2008; Zisserson, Palfai, & Saitz, 2007).

* Corresponding author. Tel.: +1 617 353 9345; fax: +1 617 353 9609.
E-mail address: palfai@bu.edu (T.P. Palfai).

0306-4603/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.addbeh.2011.01.005

Previous work has suggested that students who are heavier drinkers may be more responsive to brief alcohol interventions (see Elliott et al., 2008; Murphy et al., 2001), as heavier drinking students may be more influenced by information provided in feedback and have greater discrepancies with perceived drinking norms. The main objective of the current study was to examine the question of whether a web-based alcohol intervention that provided personalized feedback about both alcohol use and alcohol-related consequences would be differentially effective for hazardous drinkers who had experienced more frequent alcohol-related consequences. A secondary objective of this study was to examine whether assessments would differentially influence the effect of the intervention. Given previous work on assessment reactivity effects (e.g., Kypri, Langley, Saunders & Cashell-Smith, 2006; Walters, Vader, Harris, & Jouriles, 2009), we sought to examine whether the influence of the intervention would be moderated by whether students completed additional assessment instruments about psychological processes related to alcohol use and change.

2. Methods

2.1. Participants

Hazardous drinking students (30% male) volunteered to participate in the present study as part of their introductory psychology class. The study was approved by the institutional review board and students provided written informed consent. One hundred and nineteen students were enrolled in the study based on screening instruments completed in the first month of the academic year as part of a "health behaviors and college life" study. Hazardous drinking students were identified as those who either (1) consumed alcohol in the past month and scored 8 or above on the Alcohol Use Disorders Identification Test (AUDIT) (Saunders, Aasland, Babor, de la Fuente, & Grant, 1993) or (2) reported two or more heavy drinking episodes in the past month. In the current study, the mean age of students was 18.6 years (SD = 1.45). Ethnic/racial minorities comprised 20.5% of the sample (5% Hispanic, 14% Asian/Pacific Islander, and 1.5% Black).

2.2. Measures

Students completed some assessments by computer and others using pencil-and-paper. Computerized alcohol assessments provided the basis for feedback delivered to students in the alcohol intervention condition. Alcohol use and consequence measures were delivered by computer while other alcohol variables were assessed using paper and pencil. Quantity of alcohol use and heavy drinking episodes (five or more drinks for males and four or more for females) in the past month were assessed with the Daily Drinking Questionnaire-Modified (Dimeff, Baer, Kivlahan, & Marlatt, 1999). Alcohol-related problems were assessed with the Young Adult Alcohol Problems Screening Test-36 (YAAPST-36; Hurlbut & Sher, 1992). In the present study, students were asked to report the presence of problems [Y/N] over the past month. Students also completed six questions about general health-behaviors (i.e., sleep, smoking, exercise, and diet).

2.3. Procedure

Students completed screening questionnaires on health behaviors during the first month of the academic year. Those who were hazardous drinkers were scheduled for another appointment within 7 days. At this baseline session, hazardous drinkers were randomized to one of 4 conditions based on the 2 Intervention × 2 Assessment factorial design. Following alcohol assessment, those in the Control condition were provided with information on health guidelines for sleep and consumption of fruits and vegetables. Those in the Intervention condition were provided with personalized feedback about norms of total consumption and heavy drinking episodes that were university and gender specific. The intervention also included norms about low frequency alcohol-related consequences (<40%) which were personalized by highlighting specific consequences identified by each student. Finally, the intervention provided information about costs and calories associated with use (e.g., Walters et al., 2007) and information about peak blood alcohol levels associated with heavy drinking episodes. Half of the participants in each condition completed an additional series of assessments designed to assess alcohol-related motivational variables. Some measures were completed prior to the computerized screening, namely the Drinking Motives Questionnaire (Cooper, 1994) and the Alcohol Outcome Expectancies Scale (Leigh & Stacy, 1993). Other measures were completed after the intervention, specifically, (1) the Readiness to Change Questionnaire (Rollnick, Heather, Gold, & Hall, 1992), (2) the Alcohol Use Discrepancy (McNally, Palfai, & Kahler, 2005), and (3) a goal striving measure on which students listed 8 personal life goals then rated the degree to which their alcohol use facilitated or inhibited each of these strivings (e.g., Cox & Klinger, 2004; Simons, Christopher, & McLaury, 2004). Upon completion, students were scheduled for follow-up appointments to complete an additional set of alcohol measures 1-month later which served as the primary dependent measures.

3. Results

3.1. Alcohol involvement at baseline

At baseline, students reported a mean of 12.14 (SD = 7.2) drinks per week, 5.3 (SD = 3.2) heavy drinking episodes in the past month, and 6.40 (SD = 3.8) alcohol-related consequences in the past month. Of the 119 students, 77% had AUDIT scores of 8 or greater. There were no differences by intervention or assessment condition at baseline on these variables.

3.2. Weekly drinking

Hierarchical linear regression analysis was used to examine whether the effects of the intervention on weekly drinking was moderated by negative consequences. Intervention and Assessment conditions and gender were all dummy coded and the alcohol consequence variable was centered prior to calculating the interaction terms (Cohen & Cohen, 1983). Gender and weekly drinking at baseline were entered in the first step, followed by Intervention, Assessment, and Consequences on the second step. Two-way and three-way interactions were entered on steps three and four respectively. No gender interaction effects were observed in these analyses.

Analyses of the influence of intervention on number of drinks per week on Step 2 showed a significant effect of intervention on follow-up alcohol consumption, beta = −.10, $p < .05$, but no main effect of Assessment, beta = −.06, $p = .24$. However, this intervention effect was qualified by an interaction with baseline alcohol-related consequences, beta = −.18, $p < .05$. This interaction, which is depicted in Fig. 1, showed that the intervention was associated with less alcohol use at follow-up than the control condition for those with high levels of alcohol consequences but not those with low levels. Simple slope analyses showed that students with high levels of alcohol-related consequences (+1SD) exhibited significantly less drinking when exposed to the intervention condition B [SE] = −3.93 [1.068], t = −2.03, $p < .05$, whereas those who had few alcohol-related consequences (−1SD) did not, B [SE] = .71 [1.01], t = .57, p = ns. The Assessment × Intervention interaction was not significant, beta = .004, $p = .94$, nor was the assessment condition interaction with consequences, beta = .04, $p = .504$.

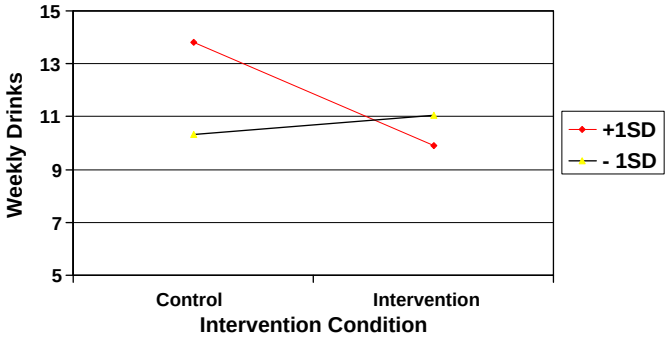


Fig. 1. Weekly total alcohol consumption at 1 month follow-up by intervention condition and baseline alcohol-related consequences.

3.3. Heavy episodic drinking

Examination of heavy drinking episodes at follow-up showed little difference between groups (Intervention M = 4.39; SD = 3.03; and Control M = 4.37; SD = 3.23). Although regression analyses did not show a main effect of intervention on heavy drinking episodes, beta = −.01, p = ns, the intervention × consequences interaction was significant, beta = −.28, p < .01. As shown in Fig. 2, the intervention appeared to influence heavy episodic drinking differentially for those with higher levels of baseline consequences. Simple slope analysis showed that students with high levels of alcohol-related consequences (+1SD) exhibited significantly fewer heavy drinking episodes when exposed to the intervention condition, B [SE] = −1.197 [.50], t = −2.29, p < .05, whereas the simple slope analysis was not significant for those with lower levels of alcohol-related consequences (−1SD), B [SE] = .93 [.51], t = 1.84, p = ns.

4. Discussion

The current study suggests that providing web-based personalized feedback about alcohol use and consequences may be a particularly effective strategy for reducing alcohol use among hazardous drinking students who have experienced high levels of alcohol-related negative consequences. Hazardous drinkers who reported high levels of alcohol-related negative consequences showed less weekly alcohol use and heavy episodic drinking if they were exposed to the alcohol intervention compared to controls. These results parallel previous work that suggests that those who experience higher levels of alcohol involvement may experience feedback as more relevant and salient (e.g., Murphy et al., 2001). Computerized feedback about norms for alcohol use and alcohol-related consequences among peers may have particular impact for students who have experienced a number of alcohol-related consequences. These students receive personally salient information that suggests that most students do not drink as much as they do and also that most students do not have the same number of negative alcohol outcomes (e.g., blackouts).

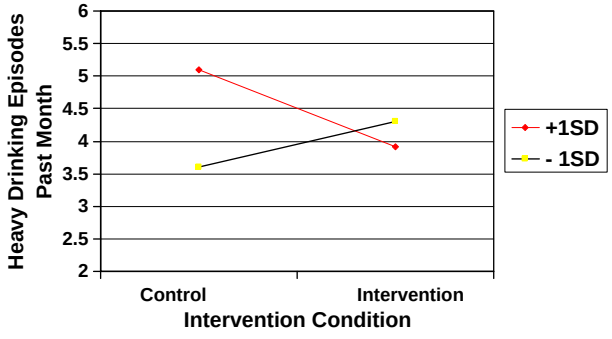


Fig. 2. Heavy drinking episodes at 1 month follow-up by intervention condition and baseline alcohol-related consequences.

Conversely, those who experienced low levels of alcohol-related negative consequences did not appear to benefit from the intervention. The findings of this study are consistent with the view that web-based interventions for hazardous alcohol use may be more efficacious if they are tailored to individual risk factors. Students may be more likely to attend to information that is relevant to personal risks which may be more likely to promote behavior change. For those who do not share those risks, however, the addition of information associated with alcohol risks may be ineffective or even detract from intervention elements. Clearly, these data cannot address this point directly as these are issues that must be explored in more detail through dismantling studies (e.g., Walters et al., 2009) to better understand how specific intervention components may interact with individual differences.

Analyses of assessment effects did not show a main effect nor was there a significant interaction between assessment and intervention. It should be noted that the comparison of assessment conditions in this study focused on the differences in the extent of motivational assessment as all participants completed multiple measures of alcohol use and problems. As a number of studies have shown that assessment may influence alcohol outcomes (e.g., Kypri et al., 2006; Walters et al., 2009) the field would clearly benefit from a more careful analysis of how the assessment of specific variables (e.g., use, problems, and psychological processes) may influence drinking outcomes and interact with the intervention effects.

In sum, web-based interventions provide a promising approach to decrease hazardous drinking among students as they expand the base of students that may be reached with screening and brief intervention approaches and provide increased flexibility for the intervention content that may be delivered. Better understanding of how intervention content may be tailored to student differences in order to maximize reductions in hazardous drinking remains a central issue in this field (Carey et al., 2007).

Role of Funding Sources

Funding for this research study was provided in part by NIAAA Grant P60 AA013759 (David Rosenbloom, PI). NIAAA had no role in the study design, collection, analysis, interpretation of data, writing of the manuscript, or decision to submit the paper for publication.

Contributors

The first author contributed to all aspects of the research including the study design, intervention development, analysis and manuscript writing. Author #2 contributed to the study design and writing of the manuscript. Author #3 contributed to the intervention development, analysis planning, and writing of the manuscript.

Conflict of Interest

There are no conflicts of interest for any of the authors.

References

Carey, K. B., Scott-Sheldon, L. A. J., Carey, M. P., & DeMartini, K. S. (2007). Individual-level interventions to reduce college student drinking: a meta-analytic review. *Addictive Behaviors*, *32*, 2469−2494.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.

Cooper, M. L. (1994). Motivations for alcohol use among adolescents: Development and validation of a four-factor model. *Psychological Assessment*, *6*, 117−128.

Cox, W. M., & Klinger, E. (2004). Systematic motivational counseling. In W. M. Cox, & E. Klinger (Eds.), *Handbook of motivational counseling: motivating people for change*. London: Wiley.

Dimeff, L. A., Baer, J. S., Kivlahan, D. R., & Marlatt, G. A. (1999). *Brief alcohol screening and intervention for college students (BASICS): A harm reduction approach.* New York: Guilford Press.

Elliott, J. C., Carey, K. B., & Bolles, J. R. (2008). Computer-based interventions for college drinking: A qualitative review. *Addictive Behaviors*, *33*, 994−1005.

Hingson, R. W., Heeren, T., Zakocs, R. C., Kopstein, A., & Wechsler, H. (2002). Magnitude of alcohol-related mortality and morbidity among US college students ages 18–24. *Journal of Studies on Alcohol*, *63*, 136−144.

Hurlbut, S. C., & Sher, K. J. (1992). Assessing alcohol problems in college students. *Journal of American College Health*, *41*, 49−58.

Kypri, K., Langley, J. D., Saunders, J. B., & Cashell-Smith, M. L. (2006). Assessment may conceal therapeutic benefit: Findings from a randomized controlled trial for hazardous drinking. *Addiction*, *102*, 62—70.

Leigh, B. C., & Stacy, A. W. (1993). Alcohol outcome expectancies: Scale construction and predictive utility in higher order confirmatory models. *Psychological Assessment*, *5*, 216—229.

McNally, A. M., Palfai, T. P., & Kahler, C. (2005). Motivational interventions for heavy drinking college students: Examining the role of psychological discrepancy processes. *Psychology of Addictive Behaviors*, *19*, 79—87.

Murphy, J. G., Duchnick, J. J., Vuchinich, R. E., Davison, J. W., Karg, R. S., Olson, A. M., et al. (2001). Relative efficacy of a brief motivational intervention for college student drinkers. *Psychology of Addictive Behaviors*, *15*, 373—379.

Neighbors, C., Larimer, M. E., & Lewis, M. A. (2004). Targeting misperceptions of descriptive drinking norms: efficacy of a computer-delivered personalized normative feedback intervention. *Journal of Consulting and Clinical Psychology*, *72*, 434—447.

Rollnick, S., Heather, N., Gold, R., & Hall, W. (1992). Development of a short readiness-to change questionnaire for use in brief opportunistic interventions among excessive drinkers. *British Journal of Addiction*, *87*, 743—754.

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption: II. *Addiction*, *88*, 791—804.

Simons, J. S., Christopher, M. S., & McLaury, A. E. (2004). Personal strivings, binge drinking, and alcohol problems. *Addictive Behaviors*, *29*, 773—779.

Walters, S. T., & Neighbors, C. (2005). Feedback intervention for college alcohol misuse: what, why, and for whom? *Addictive Behaviors*, *30*, 1168—1182.

Walters, S. T., Vader, A. M., & Harris, T. R. (2007). A controlled trial of web-based feedback for heavy drinking college students. *Prevention Science*, *8*, 83—88.

Walters, S. T., Vader, A. M., Harris, T. R., & Jouriles, E. N. (2009). Reactivity to alcohol assessment measures: an experimental test. *Addiction*, *104*, 1305—1310.

Wechsler, H., Lee, J. E., Kuo, M., & Lee, H. (2000). College binge drinking in the 1990s: A continuing problem: Results of the Harvard School of Public Health 1999 College Alcohol Study. *Journal of American College Health*, *48*, 199—210.

Zisserson, R. N., Palfai, T. P., & Saitz, R. (2007). No-contact interventions for unhealthy college drinking: efficacy of alternatives to person-delivered approaches. *Substance Abuse*, *28*, 119—131.

Patient Navigation to Increase Mammography Screening Among Inner City Women

Christine E. Phillips, MD, Jessica D. Rothstein, BA, Kristine Beaver, MPH, Bonnie J. Sherman, PhD, Karen M. Freund, MD, MPH, and Tracy A. Battaglia, MD, MPH

Women's Health Unit, Section of General Internal Medicine, Department of Medicine, and Women's Health Interdisciplinary Research Center, Boston University School of Medicine, Boston, MA, USA.

BACKGROUND: Lower mammography screening rates among minority and low income women contribute to increased morbidity and mortality from breast cancer.
OBJECTIVE: To evaluate the effect of a patient navigation intervention on adherence rates to biennial screening mammography among women engaged in primary care at an inner-city academic medical center.
DESIGN: Quality improvement intervention with a concurrent control group, conducted from February to November of 2008.
STUDY SUBJECTS: All women in a hospital-based primary care practice aged 51–70 years. Subjects were randomized at the level of their primary care provider, such that half of the patients in the practice received the intervention, while the other half received usual care.
INTERVENTIONS: Intervention subjects whose last mammogram was >18 months prior received a combination of telephone calls and reminder letters from patient navigators trained to identify barriers to care. Navigators were integrated into primary care teams and interacted directly with patients, providers, and radiology to coordinate care. Navigators utilized an electronic report to track subjects. Adherence rates to biennial mammography were assessed in intervention and control groups at baseline and post-intervention.
KEY RESULTS: A total of 3,895 women were randomized to intervention (n=1,817) and control (n=2,078) groups. Mean age was 60, 71% were racial/ethnic minorities, 23% were non-English speaking, and 63% had public or no health insurance. At baseline, there was no difference in mammography adherence between the control and intervention groups (78%, respectively, p=0.55). After the 9-month intervention, mammogram adherence was higher in the intervention group compared with the control group (87% vs. 76%, respectively, p<0.001). Except among Hispanic women who demonstrated high rates in both the intervention and control groups (85% and 83%, respectively), all racial/ethnic and insurance groups demonstrated higher adherence in the intervention group.
CONCLUSIONS: Patient navigation improves biennial mammography rates for inner city, low income, minority populations.

KEY WORDS: mammography screening; patient navigation; quality improvement; disparities; women's health.
J Gen Intern Med 26(2):123–9
DOI: 10.1007/s11606-010-1527-2

Electronic supplementary material *The online version of this article (doi:10.1007/s11606-010-1527-2) contains supplementary material, which is available to authorized users.*

Received April 28, 2010
Revised August 22, 2010
Accepted September 1, 2010
Published online October 8, 2010

INTRODUCTION

Breast cancer remains the second leading cause of cancer-related death in women; in 2009, breast cancer was responsible for an estimated 40,170 deaths in the United States[1]. While advances in early diagnosis and treatment contribute to the overall decline in breast cancer mortality, certain vulnerable populations, including racial/ethnic minorities and the poor or uninsured, are burdened with a disproportionate share of this mortality[2]. The etiology of these disparate outcomes is complex, yet delays in diagnosis and treatment have been demonstrated to play a role[3–6]. Breast cancer screening rates among medically underserved populations, including the uninsured [7] and non-Whites[3], remain substantially lower than among insured white populations.

In an effort to reduce these cancer disparities, interventions to improve mammography utilization have been tested in diverse settings, with varying success. The most effective programs have incorporated multiple strategies that target individual and system barriers[8]. Patient navigation is emerging nationally as a culturally tailored, system-based intervention that targets individual barriers in an effort to reduce cancer health disparities[9]. A growing body of literature has documented the success of navigation after an abnormal screening test is identified[10–14], but the evidence for improving mammography utilization is limited to non-generalizable target populations[15,16] or lack of rigorous control groups[17]. The purpose of this study was to examine whether a quality improvement patient navigation program could improve adherence to biennial mammography screening in a safety-net practice that

serves a largely minority, inner-city, and underinsured patient population.

METHODS

We implemented a quality improvement patient navigation intervention in 2008 to improve mammography utilization as defined by HEDIS (Healthcare Effectiveness and Data Information Set) criteria[18] among patients served in the three internal medicine practices of an academic safety-net hospital in Boston. Due to limited resources of navigators' time and the need to address a large population, navigation could not be offered to all eligible patients at the time of program implementation. In order to equitably provide services and provide an equivalent comparison group for our evaluation, we randomized at the level of the provider to initially target half the population, since randomization at the practice level may be imbalanced because of system and provider characteristics. At the end of 9 months, the concurrent control group also received the navigation services. We report here findings from the initial 9-month intervention (February–November 2008) period. The Boston Medical Center Institutional Review Board approved this study.

Study Population

Eligibility was based on HEDIS criteria and included women aged 51 to 70 who were assigned a primary care provider and had a documented visit with that provider in the previous 2 years. Women were excluded if they had documentation of bilateral mastectomy. Prior to randomization, providers were stratified into high (≥50%) versus low (<50%) HEDIS scores to balance the intervention and control groups. At baseline, provider HEDIS scores had a median value of 76% (interquartile range of 68%–82%). Half of each group of practicing providers was then randomly selected to have their eligible patients enrolled into navigation. Patients of four providers were excluded because they had five or fewer eligible patients indicating that they were not actively practicing primary care at the initiation of the study.

Patient Navigation Intervention

Navigation services targeted only women in the intervention group whose last documented mammogram was more than 18 months prior at any point during the intervention period. Three patient navigators were hired based on experience providing navigation services[10] for diverse, inner-city patients and knowledge of existing local health systems. Each completed national[19] and local navigation training programs[20] that emphasize barriers-focused culturally tailored services based upon the care management model[21]. Two were bilingual such that one spoke fluent Spanish and another fluent Portuguese and Cape Verdean Creole in addition to English.

The navigation protocol[21] included four main activities: (1) use of an electronic medical record (EMR)-based tracking system to identify eligible women, (2) identifying and (3) helping overcome individual barriers to care, and (4) tracking women through completion of mammograms. Navigators were incorporated into the primary care team in that they had regular contact and interaction with the provider about specific patient care issues. Prior to initiating navigation services, the navigator reviewed eligible patients with each provider to identify known barriers to care or recommendations from providers to not initiate navigation due to comorbidities that made screening undesirable (e.g., terminal illness from another condition). Navigators completed a series of at least three outreach telephone call attempts over a 2-week period (during daytime and early evening), followed by two letters if no contact was made, the last registered, to inform women of their need for a mammogram and the availability of the navigator to support them (see Appendix online for a detailed navigation protocol flow sheet). Upon contact, navigators inquired about individual barriers to accessing care—including but not limited to transportation issues, work scheduling conflicts, and fear—and then utilized available resources to address those barriers. The language line was available for interpreter support for non-English-speaking women if the navigator did not speak their native language. Navigators were granted scheduling access to radiology to schedule a mammogram directly.

An electronic system was developed for navigators to track patients over time. The tracking system organized women by time since the last mammogram, easily demonstrated how many telephone calls or letters had been completed as part of the protocol, and highlighted the next step needed for the protocol. When the navigator completed the entire protocol, the reason for non-adherence was documented, the provider was notified, and that subject was removed from the tracking list. Once a subject completed the mammogram, results were documented, and navigation was ended if the results were normal. Tracking of abnormal results was continued until a diagnosis was reached and the provider was aware of the results.

Data Collection

Socio-demographic data were collected directly from the electronic administrative database SDK® (Software Development Kit) and included age, race/ethnicity, health insurance coverage, primary language, marital status, and education level. Race/ethnicity was collected using a single question asked by registration clerks and has previously been shown to correlate with patient self-report in a similar population[22]. The database collapsed Portuguese Creole and Cape Verdean Creole into one category, and we report these two languages together with Portuguese in order to account for language concordance between the navigator and the patients. All commercial and employer-based insurance plans were coded as private insurance. Public insurance consisted of Medicaid, Medicare, or Commonwealth Care, the Massachusetts subsidized health insurance plan which began in November 2006[23]. Uninsured patients were covered primarily through the Massachusetts uncompensated care pool or the Centers for Disease Control breast cancer screening program[24]. We collapsed public insurance and uninsured into one category because both groups received coverage for screening mammography.

Clinical data were obtained electronically from the medical records in the Centricity® EMR. Using ICD-9 billing codes in the past 24 months, we calculated the Charlson comorbidity score[25] using the Deyo method[26]. Completion of screening mammography was determined by electronic query of the EMRs that search patient charts for evidence of an internal radiology report or outside films received. The HEDIS criterion[18] for mammography adherence was defined as completion of a bilateral screening mammogram in the past 24 months.

Navigators documented all activities in an electronic template within the EMR. These logs included the number of encounters and type of patient contacts and reason for non-adherence for those with outstanding mammograms (i.e., unable to contact, moved or transferred care, declined services, comorbidities that made screening undesirable, insurance issues, and did not keep appointment on two or more occasions).

Data Analysis

Descriptive statistics on patient socio-demographics were performed for all eligible subjects in the intervention and control arm. Statistical differences were identified using the chi-square test or t-test.

Unadjusted rates of adherence to biennial screening mammography were compared for the intervention and control groups at baseline and post intervention time periods. Unadjusted logistic regressions were performed for each demographic subgroup for the post intervention groups, while adjusted logistic regressions were performed for both time period groups. Regressions modeled adherence to biennial screening mammography (bivariate), and to control for the influence (clustering effect) of each provider on the association between the outcome and intervention group, models were performed using GEE (generalized estimation equation) logistic regression controlling for clustering on the provider level. Adjusted models controlled for all socio-demographic variables.

Adherence rates for intervention and control groups by the interval from their last mammogram at study initiation were computed along with adjusted models for each interval and group, also controlling for clustering on the provider level.

All tests were two-tailed, with a statistical significance level set at p=0.05. Individual regressions were performed for each socio-demographic variable in order to assess the benefit of the intervention for specific subgroups of the sample. All data were analyzed using Statistical Analysis System version 9.1 (SAS Institute, Cary, NC).

RESULTS

A total of 3,895 women were included in the study (1,817 intervention, 2,078 control). Table 1 shows baseline characteristics by control and intervention groups. The average age in the total population was 60 years (SD 5 years). Most women were from racial/ethnic minority groups (47% African American, 11% Hispanic). Primary language was English for most (77%), while 9% spoke a non-English language also spoken by the navigator (Spanish, Portuguese, and Cape Verdean Creole) and 14% required a interpreter services support. The majority had a public form of insurance and low educational attainment, with 7% never attending school and 34% not completing high school. Most (64%) were not married, and 34% had a Charlson comorbidity score[25] of one or greater. There was a greater percentage of Hispanic and Spanish-speaking women in the control group compared with the intervention group, reflecting that Hispanics are clustered with specific provider panels and not randomly distributed in the practice. There were two Spanish-speaking providers, one randomized to the intervention and the one to the control group.

Table 2 shows the unadjusted adherence rates and the adjusted odds ratios for mammography adherence between intervention and control patients at baseline and post intervention. At baseline, adherence rates were the same for the intervention and control groups, 78% respectively. By the end of the 9-month intervention, 87% of patients in the intervention group demonstrated biennial mammography adherence compared to 76% in the control group. After adjusting for all socio-demographic variables and using cluster analysis to adjust for provider, the odds of adherence in the intervention group was 2.5 (95% CI, 1.9–3.2) compared with the control group. Table 3 shows the odds ratios, across each demographic subgroup, for mammography adherence between intervention and control patients while clustering on the provider level, demonstrating that the intervention had a positive impact in all subgroups categorized by age, education level, marital status, insurance type, and level of comorbidity. The same was true for each racial/ethnic and language group with the exception of Hispanic women who demonstrated high baseline adherence rates in the intervention and control groups (85% and 83%, respectively).

By design, only women whose last mammogram had occurred more than 18 months ago received navigation services. Table 4 describes adherence rates for intervention and control patients by the interval from their last mammogram at study initiation in an effort to demonstrate the effect of navigation on the targeted population. This breakdown shows that the greatest improvement occurred in the two groups (18–24 months; >24 months) with the longest gaps since their previous mammography. Of those with mammograms more than 24 months before the start of the intervention, navigation adherence was 50% compared with only 17% in the control group. Navigated patients whose last mammogram had been performed more than 18 months but less than 24 months prior to the beginning of the intervention had an adherence rate of 74% compared with 37% of those in the control group. Women in the 12 to 18 month group are included here because a subset of them became eligible for navigation at some point during the study period, and indeed the intervention group showed a greater rate of adherence (97%) compared with the control group (93%).

Of the 1,817 women in the intervention group, 661 received navigation services with a mean of two telephone calls and one letter per subject. This resulted in 271 scheduled and 251 completed mammograms. Of these mammograms, 6% were abnormal [Breast Imaging Reporting and Data System (BIRADS) 0, 3, 4, or 5]. No cancers were identified during the 9-month study period. Of the patients randomized to the navigation group who remained non-adherent at the conclu-

Table 1. Baseline Characteristics of Patient Navigation Control and Intervention Subjects

Characteristic		Total (N=3,895)	Control (N=2,078)	Intervention (N=1,817)	P-value*
Age, mean (SD†), years		60 (5)	60 (5)	60 (5)	0.05
		N (%)	N (%)	N (%)	
Race	White	1,123 (29)	618 (30)	505 (28)	<0.001
	African-American	1,848 (47)	926 (45)	922 (51)	
	Hispanic	430 (11)	295 (14)	135 (7)	
	Other	494 (13)	239 (12)	255 (14)	
Insurance	Private insurance	1,432 (37)	758 (36)	674 (37)	0.69
	Public insurance	2,463 (63)	1,320 (64)	1,143 (63)	
Language	English	3,005 (77)	1,546 (74)	1,459 (80)	<0.001
	Spanish	281 (7)	204 (10)	77 (4)	
	Portuguese/Cape Verdean Creole	60 (2)	29 (1)	31 (2)	
	Other	549 (14)	299 (14)	250 (14)	
Education level	Did not attend school	286 (7)	160 (8)	126 (7)	0.003
	Did not graduate high school	1,340 (34)	688 (33)	652 (36)	
	Graduated high school or GED‡	840 (22)	469 (23)	371 (20)	
	Some college/voc/tech program	684 (18)	396 (19)	288 (16)	
	Graduated college/post grad	580 (15)	282 (14)	298 (16)	
Marital status	Married	1,409 (36)	738 (36)	671 (37)	0.31
Level of comorbidities§	Score 0	2,569 (66)	1,395 (67)	1,174 (65)	0.18
	Score1	997 (26)	517 (25)	480 (26)	
	Score 2+	317 (8)	162 (8)	155 (9)	

*P-value is based on chi-square test, with a statistical significance level set at p=0.05
†SD denotes standard deviation
‡GED denotes General Education Diploma
§*Comorbidity assessed using the Charlson Index[25], using the Deyo methodology*

sion of the study, 61% could not be contacted despite multiple telephone calls and letters, 14% moved or transferred care to another facility, and 14% declined mammography.

DISCUSSION

Using a comparable continuous control group, our study demonstrates the impact of a patient navigation program in primary care at achieving expected mammography screening rates in a diverse inner-city underserved population. With the exception of Spanish speakers and the Hispanic population, who at baseline had high rates, the navigation intervention increased adherence across all ages, insurance groups, education levels and all other languages and races. Our study design, evaluating a quality improvement navigation program in an entire practice of vulnerable patients, demonstrates the feasibility of adopting this method of care to a clinical setting that mirrors urban safety net settings throughout the country.

Table 2. Adjusted Patient Navigation Intervention Affects on Mammography Adherence Controlling for Baseline Characteristics

	Control (N=2,078)	Intervention (N=1,817)	
	Adherent (n%)*		Adjusted OR (95% CI)†
Baseline	1,631 (78)	1,412 (78)	1.1 (0.8–1.4)
Post Intervention	1,589 (76)	1,575 (87)	2.5 (1.9–3.2)

*Unadjusted mammography adherence frequencies
†Odds ratios from a logistic regression analysis, adjusted for baseline characteristics variables and clustering on the provider level

Our findings of improved mammogram utilization with a patient navigation intervention are consistent with existing literature[15–17]. Dignan and colleagues[15] showed improved adherence to mammogram screening over an 18-month period in a randomized controlled trial (RCT) of 157 Native American women utilizing direct patient contact or telephone calls versus a control group. Paskett and colleagues' RCT[16] of 851 subjects also showed improved adherence of mammogram screening in a rural, low income population of white, African American and Native American subjects. Han and colleagues' study[17] utilized patient navigation to improve mammogram adherence of 102 Korean American women after 6 months, but lacked a control group for comparison. Our study findings, designed as a practice improvement within an urban safety-net setting, included a more diverse, yet vulnerable population and thus provide further evidence for the generalizability of navigation as a means to reduce cancer health disparities.

At baseline, more Hispanic and Spanish-speaking women were present in the group allocated to the control group, reflecting that patients are not randomly distributed throughout the practice, but rather are more likely to be seen by Spanish-speaking providers. Prior research has shown both lower[27,28] and higher[29,30] mammography screening rates among Hispanic as compared with white women and may reflect a variety of local factors, including the community's overall acculturation and education levels, as well as access to insurance and to bilingual health care providers.

Our study was developed to improve HEDIS rates as a quality improvement project within primary care and thus

Table 3. Unadjusted Patient Navigation Intervention Affects on Mammography Adherence Stratified by Baseline Characteristics

		Control (N=2,078)	Intervention (N=1,817)	
		Adherent (n%)*		Stratified OR (95% CI)†
Post intervention by characteristic				
Age	50–59 years	743 (74)	835 (86)	2.2 (1.6–2.9)
	60–70 years	846 (79)	740 (87)	1.9 (1.3–2.6)
Race	White	433 (70)	429 (85)	2.4 (1.5–4.0)
	African American	724 (78)	806 (87)	1.9 (1.4–2.6)
	Hispanic	245 (83)	115 (85)	1.2 (0.8–1.8)
	Other	187 (78)	225 (88)	2.1 (1.3–3.3)
Insurance	Public	1,002 (76)	986 (86)	2.0 (1.6–2.6)
	Private	587 (77)	589 (87)	2.0 (1.3–3.2)
Language	English	1,172 (76)	1,285 (88)	2.4 (1.7–3.3)
	Spanish	167 (82)	64 (83)	1.1 (0.6–1.9)
	Portuguese/Cape Verdean Creole	21 (72)	28 (90)	3.6 (1.2–10.3)
	Other	229 (77)	198 (79)	1.2 (0.8–1.7)
Education level	Did not attend school	115 (72)	113 (90)	3.4 (1.7–6.8)
	Did not graduate high school	536 (78)	568 (87)	1.9 (1.3–2.8)
	Graduated high school or GED	401 (86)	344 (93)	2.2 (1.3–3.7)
	Some college/voc/tech program	299 (76)	258 (90)	2.8 (1.9–4.1)
	Graduated college/post grad	183 (65)	226 (76)	1.7 (1.1–2.5)
Marital status	Married	570 (77)	596 (89)	2.1 (1.6–2.9)
	Not married	1,019 (76)	979 (87)	2.3 (1.5–3.6)
Level of comorbidities	Score 0	1,052 (75)	1,019 (87)	2.1 (1.5–3.0)
	Score 1	408 (79)	417 (87)	1.8 (1.3–2.4)
	Score 2+	125 (77)	132 (85)	1.9 (1.1–3.3)

*Unadjusted mammography adherence frequencies
†Unadjusted odds ratios from a logistic regression analysis clustering on the provider level

performed as a population-level analysis. The study design responded to a growing emphasis on HEDIS rates as a quality performance measure of individual providers. Interventions like this patient navigation program are becoming increasingly important in order to ensure that practices serving populations with historically lower screening rates achieve benchmarks, both for patient care and for practice reimbursement under pay-for-performance plans. One strength of our study is that it was integrated into the practice with provider "buy-in," and was designed to evaluate the benefit and effectiveness of integration of this type of program into a busy primary care practice. As such, this model of care fits the Medical Home Model, which is increasingly recognized as a standard way to transform care delivery in primary care settings[31–34].

We designed our intervention to improve our practice HEDIS measures of biennial screening for all women 51–70 years of age, at a time when the providers were recommending annual screening. Best practice guidelines for patient navigation and timing of the intervention have yet to be defined as reflected by the different intervention protocols implemented in prior studies. Paskett and colleagues[16] navigated those ≥12 months overdue for a mammogram and followed them for an additional 12–14 months, while Han and colleagues[17] navigated women 2 years overdue and followed up at 6 months. Dignan and colleagues[15] navigated for patients 18 months overdue. We chose to initiate navigation after an 18-month screening interval and found this to be an effective strategy in the rational use of the navigator resources, as evidenced by the fact that our greatest effect of the intervention was seen among the groups with the longest time interval since their last mammogram. Even with changes in recent guidelines, our protocol is consistent with recommended mammography frequency in this age group[27,35–37]. There is still a subset of women who remain non-adherent to mammography screening using this protocol; the majority of these women were not reachable by phone or mail based on available contact information. This reflects

Table 4. Mammography Adherence by Time Since Last Mammogram for Control and Intervention Subjects with Adjusted Patient Navigation Intervention Odds Ratios

	Control		Intervention		
Time since last mammogram at baseline	N	Subjects adherent post intervention	N	Subjects adherent post intervention	Adjusted OR (95% CI)*
>24	447	78 (17)	405	202 (50)	5.6(3.9–8.2)
18–24	126	47 (37)	89	66 (74)	6.0(2.8–12.7)
12–18	626	585 (93)	547	531 (97)	3.5 (1.8–6.5)

*Odds ratios from a logistic regression analysis, adjusted for baseline characteristics variables, and clustering on the provider level

communication challenges in caring for an inner city, at-risk population and suggests a different approach is necessary for this group.

Our study is somewhat limited in its generalizability because it was conducted in only three practices at a single academic safety net institution and required the use of an EMR and information technology support. However, our system reflects standard EMR support and practice systems that have become federal mandates[38,39]. Due to limited resources, we were unable to assess costs of the program or patient or provider satisfaction, all of which are crucial to sustainability of such programs.

Our findings support the benefit of patient navigation programs in the primary care setting as one approach to reduce cancer health disparities. While financial support is necessary for primary care providers to develop and maintain such programs, the Medical Home Model[31–34] could be one venue to provide the infrastructure and personnel necessary for sustainable navigation implementation. Health care policy-makers should continue to explore advocacy efforts in order to determine how to sustain these programs.

Acknowledgements: *Funded by the Avon Foundation Safety Net Grant. The study would like to acknowledge the contributions of Faber Alvis, Mariuca Tuxbury, and Arlene Ash, PhD. This study was presented as an oral presentation at the 32nd annual meeting of the Society of General Internal Medicine conference in Miami in May 2009.*

Conflicts of Interest: *None disclosed.*

Corresponding Author: *Christine E. Phillips, MD, , 801 Massachusetts Ave., Crosstown 2nd floor, Boston, MA 02118, USA (e-mail: chphilli@bmc.org).*

REFERENCES

1. **National Cancer Institute.** Breast Cancer. <http://www.cancer.gov/cancertopics/types/breast. Accessed September 13, 2010.
2. **Centers for Disease Control and Prevention.** Breast Cancer. <http://www.cdc.gov/cancer/breast/statistics/race.htm >. Accessed September 13, 2010.
3. **Kagawa-Singer M, Dadia AV, Yu MC, Surbone A.** Cancer, culture, and health disparities: time to chart a new course? CA Cancer J Clin. 2010;60(1):12–39.
4. **Hershman D, McBride R, Jacobson JS, et al.** Racial disparities in treatment and survival among women with early-stage breast cancer. J Clin Oncol. 2005;23(27):6639–46.
5. **Harper S, Lynch J, Meersman SC, Breen N, Davis WW, Reichman MC.** Trends in area-socioeconomic and race-ethnic disparities in breast cancer incidence, stage at diagnosis, screening, mortality, and survival among women ages 50 years and over (1987–2005). Cancer Epidemiol Biomark Prev. 2009;18(1):121–31.
6. **Smith-Bindman R, Miglioretti DL, Lurie N, et al.** Does utilization of screening mammography explain racial and ethnic differences in breast cancer? Ann Intern Med. 2006;144(8):541–53.
7. **Centers for Disease Control and Prevention.** Behavioral Risk Factor Surveillance System summary data quality reports. Atlanta, GA: Department of Health and Human Services. <www.cdc.gov/brfss/technical_infodata/quality.htm>. Accessed September 13, 2010.
8. **Legler J, Meissner HI, Coyne C, Breen N, Chollette V, Rimer BK.** The effectiveness of interventions to promote mammography among women with historically lower rates of screening. Cancer Epidemiol Biomark Prev. 2002;11(1):59–71.
9. **Wells KJ, Battaglia TA, Dudley DJ, et al.** Patient navigation: state of the art or is it science? Cancer. 2008;113(8):1999–2010.
10. **Battaglia TA, Roloff K, Posner MA, Freund KM.** Improving follow-up to abnormal breast cancer screening in an urban population. A patient navigation intervention. Cancer. 2007;109(2 Suppl):359–67.
11. **Ell K, Vourlekis B, Lee PJ, Xie B.** Patient navigation and case management following an abnormal mammogram: a randomized clinical trial. Prev Med. 2007;44(1):26–33.
12. **Psooy BJ, Schreuer D, Borgaonkar J, Caines JS.** Patient navigation: improving timeliness in the diagnosis of breast abnormalities. Can Assoc Radiol J. 2004;55(3):145–50.
13. **Ferrante JM, Chen PH, Kim S.** The effect of patient navigation on time to diagnosis, anxiety, and satisfaction in urban minority women with abnormal mammograms: a randomized controlled trial. J Urban Health. 2008;85(1):114–24.
14. **Lobb R, Allen JD, Emmons KM, Ayanian JZ.** Timely care after an abnormal mammogram among low-income women in a public breast cancer screening program. Arch Intern Med. 2010;170(6):521–8.
15. **Dignan MB, Burhansstipanov L, Hariton J, et al.** A comparison of two Native American Navigator formats: face-to-face and telephone. Cancer Control. 2005;12(Suppl 2):28–33.
16. **Paskett E, Tatum C, Rushing J, et al.** Randomized trial of an intervention to improve mammography utilization among a triracial rural population of women. J Natl Cancer Inst. 2006;98(17):1226–37.
17. **Han HR, Lee H, Kim MT, Kim KB.** Tailored lay health worker intervention improves breast cancer screening outcomes in non-adherent Korean-American women. Health Educ Res. 2009;24(2):318–29.
18. **HEDIS.** Technical Specifications, Vol. 2. Washington, D.C: National Committee for Quality Assurance; 2007.
19. **Calhoun EA, Whitley EM, Esparza A, et al.** A National Patient Navigator Training Program. Health Promot Pract. 2008. doi:10.1177/1524839908323521. Accessed September 13, 2010.
20. Central Massachusetts Area Health Education Center: Outreach Worker Training Institute. <http://cmahec.org/index.php?option=com_content&view=article&id=46&Itemid=59>. Accessed September 13, 2010.
21. **Ward MD, Rieve JA.** The role of case management in disease management. In: **Todd WE, Nash E, eds.** Disease management: A Systems Approach to Improving Patient Outcomes. Chicago, IL: American Hospital Publishing; 1997:235–59.
22. **Laiteerapong N, Sherman BJ, Kim K, Freund K.** Validation of racial categorization in a hospital administrative database. J Gen Intern Med. 2005;20(Suppl 1):130.
23. Health Connector Facts and Figures, March 2009. Commonwealth Connector. <http://www.mahealthconnector.org/portal/site/connector/menuitem.d7b34e88a23468a2dbef6f47d7468a0c?fiShown=default>. Accessed September 13, 2010.
24. **Centers for Disease Control and Prevention.** National Breast and Cervical Cancer Early Detection Program Fact Sheet 2004/2005. Atlanta, GA: Centers for Disease Control and Prevention; 2005.
25. **Charlson ME, Pompei P, Ales KL, MacKenzie CR.** A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373–83.
26. **Deyo RA, Cherkin DC, Ciol MA.** Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol. 1992;45(6):613–9.
27. **Smith RA, Cokkinides V, Brooks D, Saslow D, Brawley OW.** Cancer screening in the United States, 2010: a review of current American Cancer Society guidelines and issues in cancer screening. CA Cancer J Clin. 2010;60(2):99–119.
28. **David MM, Ko L, Prudent N, Green EH, Posner MA, Freund KM.** Mammography use. J Natl Med Assoc. 2005;97(2):253–61.
29. **Palmer RC, Fernandez ME, Tortolero-Luna G, Gonzales A, Mullen DP.** Acculturation and mammography screening among Hispanic women living in farmworker communities. Cancer Control. 2005;12(Suppl 2):21–7.
30. **Zambrana RE, Breen N, Fox SA, Gutierrez-Mohamed ML.** Use of cancer screening practices by Hispanic women: analyses by subgroup. Prev Med. 1999;29(6 Pt 1):466–77.
31. **National Committee for Quality Assurance.** Physician Practice Connections–Patient-Centered Medical Home (PPC-PCMH). <http://www.ncqa.org/tabid/631/Default.aspx>. Accessed September 13, 2010.
32. **Fisher ES.** Building a medical neighborhood for the medical home. N Engl J Med. 2008;359(12):1202–5.
33. **Understanding the Patient-Centered Medical Home.** Philadelphia: American College of Physicians, 2010. <http://www.acponline.org/advocacy/where_we_stand/medical_home/. Accessed September 13, 2010.

34. Joint Principles of the Patient-Centered Medical Home Released by Organizations Representing More Than 300,000 Physicians. Philadelphia: American College of Physicians, 2007. <http://www.acponline.org/pressroom/pcmh.htm. Accessed September 13, 2010.
35. Screening for breast cancer: US Preventive Services Task Force recommendation statement. Ann Intern Med. 2009;151(10):716–26
36. **Qaseem A, Snow V, Sherif K, Aronson M, Weiss KB, Owens DK.** Screening mammography for women 40 to 49 years of age: a clinical practice guideline from the American College of Physicians. Ann Intern Med. 2007;146(7):511–5.
37. **The American Congress of Obstetricians and Gynecologists.** ACOG Responds to the U.S. Preventive Services Task Force's Breast Cancer Guidelines. <http://www.acog.org/acog_districts/dist_notice.cfm?recno=13&bulletin=3160. Accessed September 13, 2010.
38. **Steinbrook R.** Health Care and the American Recovery and Reinvestment Act. N Engl J Med. 2009;360(11):1057–1060.
39. **U.S. Department of Health and Human Services.** Regulations and Guidance. <http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov_regulations_and_%20guidance/1496>. Accessed September 13, 2010.

Outcomes registry for better informed treatment of atrial fibrillation: Rationale and design of ORBIT-AF

Jonathan P. Piccini, MD, MHS,[a] Elizabeth S. Fraulo, BSN,[a] Jack E. Ansell, MD,[b] Gregg C. Fonarow, MD,[c] Bernard J. Gersh, MB, ChB, DPhil,[d] Alan S. Go, MD,[e] Elaine M. Hylek, MD, MPH,[f] Peter R. Kowey, MD,[g] Kenneth W. Mahaffey, MD,[a] Laine E. Thomas, PhD,[a] Melissa H. Kong, MD,[a] Renato D. Lopes, MD, PhD, MHS,[a] Roger M. Mills, MD,[h] and Eric D. Peterson, MD, MPH[a,i] *Durham, NC; New York, NY; Boston, MA; Los Angeles, and Oakland, CA; Rochester, MN; Wynnewood, PA; and Raritan, NJ*

Background Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia and is associated with an increased risk of stroke, heart failure, and death. Data on contemporary treatment patterns and outcomes associated with AF in clinical practice are limited.

Methods/Design The Outcomes Registry for Better Informed Treatment of Atrial Fibrillation is a multicenter, prospective, ambulatory-based registry of incident and prevalent AF. The registry will be a nationwide collaboration of health care providers, including internists, primary care physicians, cardiologists, and electrophysiologists. Initial target enrollment is approximately 10,000 patients to be recruited from approximately 200 US outpatient practices. Enrolled patients will be observed for ≥2 years. A patient-reported outcomes substudy in ≥1,500 patients will provide serial quality-of-life assessments. The goal is to characterize treatment and outcomes of patients with AF, thereby promoting better quality of AF care and improved patient outcomes.

Conclusion The Outcomes Registry for Better Informed Treatment of Atrial Fibrillation will provide insights into "real-world" treatment including rate and rhythm control, stroke prevention, transitions to new therapies, and clinical and patient-centered outcomes among patients with AF in community practice settings (ClinicalTrials.gov NCT01165710). (Am Heart J 2011;162:606-612.e1.)

Atrial fibrillation (AF) is the most common sustained arrhythmia in the United States. By 2020, an estimated 7 million Americans will have AF.[1] Atrial fibrillation is associated with significant morbidity and mortality[2-4]; in particular, it increases the risk for stroke by 2- to 5-fold[5] and impairs quality of life in many patients. In economic terms, AF has a significant impact on the US health care system; treatment of newly diagnosed AF costs the Medicare program >$15 billion annually.[6]

New therapies for both stroke prevention and rhythm control are emerging from phase III trials, but data on their clinical effectiveness and safety outside clinical trial settings are lacking.[7-9] Randomized, controlled trials include selected populations, often enrolling patients with a lower burden of comorbid disease and excluding those with severe or end-stage comorbidities; thus, results of these trials may not translate well into contemporary clinical practice populations.[10,11] In addition, clinical trial participants are often treated by disease experts using standardized study protocols with frequent monitoring and closer follow-up than would be observed in routine clinical practice.

The Outcomes Registry for Better Informed Treatment of Atrial Fibrillation (ORBIT-AF) is a prospective, multicenter, outpatient registry that will identify treatment patterns of AF and variation in contemporary clinical practice according to patient demographics, clinical factors, risk stratification, provider specialty, and geographic region. Patient health status, health resource use, and clinical outcomes will be collected for ≥2 years of follow-up. This registry will focus on the adoption, use, effectiveness, and safety of novel and emerging treatments for AF, including antithrombotic therapy for stroke prophylaxis. The registry is also designed to identify reasons and predictors for lack of anticoagulation in appropriate candidates as well as noncompliance, quality

From the [a]Duke Clinical Research Institute, Durham, NC, [b]Department of Medicine, Lenox Hill Hospital, New York, NY, [c]Division of Cardiology, University of California, Los Angeles, CA, [d]Department of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, [e]Division of Research, Kaiser Permanente Northern California, Oakland, CA, [f]Department of Medicine, Boston University School of Medicine, Boston, MA, [g]Lankenau Medical Center and Institute for Medical Research, Main Line Health Systems, Wynnewood, PA, and [h]Ortho-McNeil Janssen Scientific Affairs, LLC, Raritan, NJ.
[i]On behalf of the ORBIT-AF Steering Committee and Investigators.
Jeanne E. Poole, MD served as guest editor for this article.
Submitted March 1, 2011; accepted July 6, 2011.
Reprint requests: Jonathan P. Piccini, MD, MHS, PO Box 17969, Durham, NC 27710.
E-mail: jonathan.piccini@duke.edu
0002-8703/$ - see front matter

doi:10.1016/j.ahj.2011.07.001

Table I. The ORBIT-AF enrollment criteria

Inclusion criteria	Exclusion criteria
Age ≥18 y	Anticipated life expectancy <6 m
Incident or prevalent AF with electrocardiographic documentation	Transient AF secondary to a reversible condition (eg, hyperthyroidism, pulmonary embolism, postcardiothoracic surgery)
Anticipated ability to adhere to regularly scheduled follow-up visits	Employee or immediate family member of the investigator/study center
Signed informed consent document indicating that patients understand the purpose and procedures of the study and are willing to participate in the registry	Current enrollment in a randomized clinical trial of antithrombotic therapy for AF

of care (time in therapeutic range), and discontinuation of therapy. Finally, ORBIT-AF will assess the adoption and use of AF therapies in an era marked by multiple, novel, emerging pharmacologic and nonpharmacologic interventions for the maintenance of sinus rhythm, including antiarrhythmic drug therapy and catheter ablation.

Methods

Registry objectives

The ORBIT-AF registry has 5 main objectives: (1) characterize and describe a large representative AF population in the United States, including demographics, comorbidities, and risk profiles; (2) define current practice patterns for the treatment of patients with AF, particularly stroke prevention therapies; (3) identify how patterns of care and subsequent outcomes vary by risk stratification (ie, low risk vs high risk), including existing and novel risk prediction schema; (4) assess adherence and resource use associated with current anticoagulant prophylaxis as recommended by current guidelines[12,13]; and (5) assess the adoption and impact of emerging antithrombotic and antiarrhythmic therapies on outcomes in AF, including patient-reported outcomes (PRO) and health care resource use.

Site selection

To ensure inclusion of a broad spectrum of representative patients with AF, a variety of outpatient practices, including internal medicine, cardiology, and electrophysiology clinics, were targeted for this registry. Although investigator specialty identification represents the most robust and reproducible method to ensure site heterogeneity, geographic diversity will also be considered in final site selection and patient enrollment. To further enhance heterogeneity, an adaptive registry design will be used, allowing for modifications to the registry after study initiation without affecting its validity and integrity.[14] Approximately 200 centers are expected to participate.

Study population

Investigators will enroll consecutive patients meeting inclusion criteria (Table I). All patients ≥18 years with electrocardiographically documented AF will be eligible. Patients with atrial flutter only will not be enrolled. Upon study entry, patients will be classified by type of AF (paroxysmal, persistent, and permanent AF/long-standing persistent), per the ACC/AHA/ESC 2006 Guidelines for the Management of Patients with AF and the 2007 HRS/EHRA/ECAS Expert Consensus Statement on Catheter and Surgical Ablation of Atrial Fibrillation.[2,15] If the type of AF cannot be classified at study entry (first-detected AF), it will be categorized after follow-up evaluations.

Data collection and follow-up

The primary data source will be the medical record of each enrolled patient. Sites will transmit data via a Web-enabled case report form. Patient-reported outcome questionnaires will be administered to a subsample of approximately 1,500 patients. For patients who consent to answer PRO questionnaires, the questionnaires will be administered at the baseline visit and during follow-up visits.

Data collection will occur at 6-month intervals for a minimum of 2 years after enrollment (Table II). Data collection windows will be wide (3 months in either direction) to maximize data collection during the patients' regularly scheduled follow-up visits. Data capture will include age, sex, race/ethnicity, insurance status, education level, cardiovascular risk factors, date of diagnosis, type of AF, pharmacologic treatment strategy (rate control vs rhythm control), AF ablation history, cardioversion history, transthoracic and transesophageal echocardiographic findings, antithrombotic therapy and monitoring (international normalized ratios [INRs]), concomitant medications, insurance status and provider information, comorbidities, and outcomes. The case report form also captures the specialty of the enrolling physician and that of each physician participating in the patient's AF-related care.

Outcomes

The primary outcome event in ORBIT-AF is stroke or non–central nervous system (CNS; systemic) systemic embolism. Consistent with recent clinical trials, stroke will be defined as a new, sudden, focal neurologic deficit that persists beyond 24 hours and is not due to a readily identifiable, nonvascular cause (eg, seizure).[16,17] Primary outcome events will be verified by single-source document submission (eg, hospital discharge report) and central review at the data coordinating center. The major safety outcome of interest will be major bleeding as defined by the International Society of Thrombosis and Haemostasis.[18]

Secondary outcomes will include *major adverse cardiac events* (MACE), defined as stroke or non-CNS systemic embolism, myocardial infarction, and cardiovascular death (online Appendix A). Additional secondary outcomes will include all-cause mortality, cardiovascular death, intracranial bleeding, myocardial infarction, sudden cardiac death, heart failure–related death, AF-related quality of life, and all-cause hospitalization (subcategorized as cardiovascular; bleeding related; and noncardiovascular, non–bleeding related). Anticoagulation-related secondary outcomes of interest will include

Table II. Time and events schedule

	Study day 1	Data collection interval 6 m	12 m	18 m	24 m
Assessments/procedures*					
Informed consent obtained	X				
Inclusion/Exclusion criteria	X				
Demographics	X				
Past medical history	X				
Vital signs	X	X	X	X	X
Rhythm and arrhythmia-related symptoms†	X	X	X	X	X
Interventions and treatments	X	X	X	X	X
Antithrombotic therapies	X	X	X	X	X
Current drug therapies	X	X	X	X	X
INR values	X	X	X	X	X
Outcomes	X	X	X	X	X
PRO questionnaires					
AFEQT	X		X		X
ACTS	X	X	X		X

* Sites will consecutively screen and enroll eligible patients whenever possible.
† Symptoms *include* palpitations, lightheadedness/dizziness, syncope/fainting, dyspnea at rest, dyspnea on exertion, fatigue, exercise intolerance, chest tightness/discomfort, and the European Heart Rhythm Association score.

anticoagulation-related treatment satisfaction and quality of anticoagulation as assessed by time in therapeutic range and primary discontinuation of oral anticoagulation. Use of AF-related procedures, including but not limited to transesophageal echocardiography, cardioversion, atrioventricular node ablation, pacemaker implantation, left atrial appendage closure, and pulmonary vein isolation, will also be recorded for analysis.

The INR collection will include all available prior INR measurements as recorded in the medical record for all patients receiving warfarin at enrollment. The INR measurements will also be captured prospectively in all patients.

Patient-reported outcomes: AF and anticoagulation-related quality of life

Given the prominence of quality of life considerations in AF treatment decisions, the Atrial Fibrillation Effect on QualiTy-of-life questionnaire (AFEQT; St Jude Medical, St Paul, MN) will be administered in the PRO cohort (n = 1,500) at study entry, at 12 months, and at 24 months. The AFEQT is a 20-item questionnaire assessing 3 domains of AF-related quality of life, including activity, symptoms, and treatment concerns.[19]

To determine the degree of satisfaction and quality of life in patients receiving antithrombotic therapy, the Anti-clot Treatment Scale (ACTS) questionnaire will be administered to PRO cohort patients taking an antithrombotic medication at study entry, at 6 months, at 12 months, and at 24 months. The ACTS questionnaire is a 17-item scale assessing the burdens (eg, bleeding risks, limitations, inconveniences) and benefits (confidence, reassurance, and satisfaction) of anticoagulation. For patients starting oral antithrombotic therapy at enrollment, the ACTS for incident patients will be completed and returned to the site 1 month after the patient starts anticoagulation treatment. The ACTS questionnaire will only be administered to patients receiving antithrombotic therapies.

Statistical analyses

Initial analyses will examine patient characteristics, pharmacoepidemiology, and quality of care, including guideline-based use of oral anticoagulation in eligible patients. Subsequently, attention will turn to the relationship between risk stratification, treatment, adherence, and outcomes. The association between primary and secondary outcomes, INR control, and discontinuation of oral anticoagulation will also be reported. Variables with <15% missing will routinely be imputed, although missing data will be addressed on an analysis-specific basis. Univariate and multivariable approaches will be used to identify factors associated with primary and secondary outcomes. Propensity score techniques will be used to balance comparison groups according to baseline factors.

The ORBIT-AF will also explore variation in care, including variation in performance measures for AF management, including stroke prophylaxis.[12] Characteristics of interest will include sex; race; age; prior stroke/transient ischemic attack (TIA); geographic region; socioeconomic status; $CHADS_2$, CHA_2DS_2-VASc, and HAS-BLED scores; management strategies (eg, rate control vs rhythm control, catheter ablation vs medical management, and combination therapy with antiplatelet agents); and provider specialties.[2,20,21] Pharmacologic and nonpharmacologic methods of maintaining sinus rhythm will also be analyzed in the registry. Comparative effectiveness and safety analyses will be conducted, including comparisons of individual antiarrhythmic drugs, comparisons of rhythm and rate control, and catheter ablation versus medical therapy.

Continuous variables will be summarized by medians with 25th and 75th percentiles, and categorical variables will be reported as counts with percentages. Associations will be reported using hazards ratios (for time-to-event analyses) or odds ratios (for logistic regression analyses) with 95% CIs. Statistical significance will be declared when a 2-sided α is <.05.

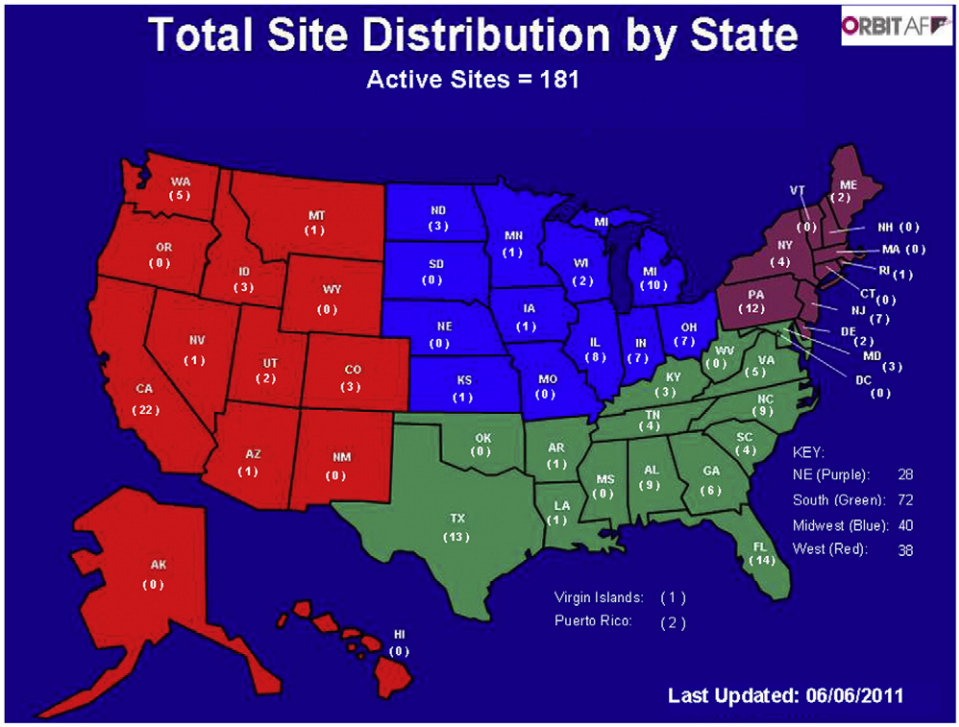
Sample size

The goal of enrolling 10,000 patients is to capture a large, representative population of patients with AF in the United States. Based on recently completed clinical trials and observational studies, we anticipate an incident stroke rate of 1.5% to 2.2% in our unselected ORBIT-AF population.[9,22,23] This would correspond to approximately 150 to 220 strokes in the first year of follow-up. We anticipate that annual warfarin discontinuation rates will range from 15% to 25% based on previous community-based studies.[24] Assuming that >50% of the cohort will be taking warfarin, we anticipate >750 therapeutic discontinuations over the first year of follow-up. This event rate will allow for even more robust model construction to identify risk factors for long-term warfarin intolerance.

Because the registry design is adaptive in nature, sample size can be modified during the course of the study to ensure adequacy of the registry to answer targeted research questions. Adaptive changes to the sample size and eligibility criteria for substudies will be considered during the late phase of patient enrollment when sufficient patient characteristic data are available to make the determination. Overall patient characteristics will be compared before and after adaptive changes to ensure that generalizability is not impaired.

The PRO substudy will compare quality of life and treatment satisfaction in 4 analytic groups (2 study groups [warfarin and

Figure 1



Geographic distribution of ORBIT-AF enrolling sites.

non-warfarin] and 2 strata [incident and prevalent]). Using the 5-point Likert scale as the basis for sample size estimation (such as the items in the ACTS), we will require 253 patients in each group to have 80% power to detect a 0.25-point difference in the mean score of an item between 2 groups (assuming an SD of unity, a 2-sided $\alpha = .05$). Accounting for the incident and prevalent strata as well as treatment discontinuation and switching, a PRO subsample of approximately 1,500 is robust for the proposed analyses.

Registry organization and funding

The study protocol was reviewed and approved by the Duke University Medical Center Institutional Review Board (IRB) and the IRB at each enrolling center. The registry is funded by Ortho-McNeil Janssen Scientific Affairs, LLC. Duke University and Duke Clinical Research Institute, along with the sponsor, will be responsible for ORBIT-AF and its design, implementation, and leadership. The Scientific Leadership Committee will include the study chair, the executive committee, and the steering committee (online Appendix B).

Results

Initial enrolling site characteristics

Given the adaptive design of the registry, the composition of the enrolling sites is continuously monitored to ensure adequate provider, demographic, and geographic diversity and representativeness. Across the United States, 161 of 181 active sites enrolled the first 7,294 patients: 58 (36%) primary care and internal medicine clinics, 86 (53%) cardiology clinics, and 17 (11%) electrophysiology clinics. At the time of this writing, 181 sites were IRB approved and activated to enroll patients. As shown in Figure 1, the sites are distributed across the United States: West (n = 38), Midwest (n = 40), South (n = 72), Northeast (n = 28), Puerto Rico (n = 2), and the Virgin Islands (n = 1).

Discussion

The Institute of Medicine has identified the management of AF as a top priority for comparative effectiveness research.[25] Atrial fibrillation increases the risk of stroke, cognitive impairment, and disability. Despite the presence of effective pharmacologic therapy for the prevention of stroke, anticoagulation prophylaxis in patients with AF is underused.[26] Because many patients with AF do not receive anticoagulation therapy or discontinue therapy within 1 year,[27] a key goal of ORBIT-AF will be to identify reasons and risk factors for nonreceipt and discontinuation of anticoagulation therapy.

Prior registry efforts

Prior registries have examined the prognosis and management of AF in clinical practice (Table III). The Euro Heart Survey on Atrial Fibrillation enrolled 5,333

Table III. Characteristics of large AF registries

	Population	Enrollment	Follow-up	Study period	No. of countries	No. of sites	Quality-of-life measures
ORBIT-AF	Outpatient	10 000	Minimum 2 y	2009-present	1 (United States)	~200	AFEQT and ACTS
Realise AF[28]	Inpatient and outpatient	10 523	Cross-sectional	2009-2010	26	831	None
RECORD AF[29]	Inpatient and outpatient	5604	1 y	2007-2008	21	532	None
AFFECTS[26]	Outpatients without significant structural heart disease	1461	1 y	2005-2007	1 (United States)	248	None
Euro Heart Survey on AF[30,31]	Inpatient and outpatient	5333	1 y	2003-2004	35	182	None
ATRIA[32]	Large integrated health care delivery system	13 559	Cross-sectional	1996-1997	1 (United States)	18	None

Registries of ≥1,000 patients. The ORBIT-AF, Realise AF, RECORD AF, and AFFECTS are formal registries that prospectively enrolled patients after obtaining informed consent.

patients with AF from 2003 to 2004.[30] At baseline, 44% of these patients were taking oral anticoagulants. The international REgistry on Cardiac rhythm disORDers (RECORD AF) enrolled 5,814 patients in 21 countries from 2007 to 2008.[29] At study entry, 55% of the cohort was treated with a rhythm control strategy, 10% had a history of stroke or TIA, and 61% were receiving oral anticoagulation. The Atrial Fibrillation Focus on Effective Clinical Treatment Strategies (AFFECTS) registry enrolled 1,165 patients and found that among eligible patients ($CHADS_2$ >1), only 69% were receiving oral anticoagulation.[26] Although these and other registries have advanced our understanding of AF, they have been limited by relatively short follow-up and the absence of formal quality-of-life and other patient-centered outcomes (Table III). The ORBIT-AF will enroll ≥10,000 patients with planned follow-up of ≥2 years. This long-term follow-up will have several important advantages, including greater ascertainment of infrequent events such as stroke and intracranial hemorrhage as well as the opportunity to assess temporal changes in treatment strategies and their association with quality of life and other important outcomes.

Importance of registry data in the evaluation of emerging therapies

During the course of this registry, we anticipate that several novel antithrombotic agents will become available for the prevention of stroke in patients with nonvalvular AF. Following the results of the RELY trial in 2009, dabigatran was approved for the prevention of stroke in patients with nonvalvular AF.[7] More recently, the ROCKET-AF trial demonstrated that rivaroxaban, a novel factor Xa inhibitor, may be a safe and effective alternative to warfarin in patients with AF and moderate to high risk of stroke.[9] Additional oral anticoagulants are currently undergoing study in phase III trials, including the factor Xa inhibitors apixaban and edoxaban.[33] These anticoagulants may improve therapeutic efficacy and safety with less monitoring and inconvenience and are expected to have a substantial effect on clinical practice. Finally, nonpharmacologic interventions for stroke prevention are also under regulatory review and may reach approval; trial results suggest that left atrial appendage occlusion may be noninferior to warfarin with an acceptable safety profile in highly selected patients.[34] The ORBIT-AF will be well positioned to monitor the diffusion of these new therapies and their safety profiles in clinical practice.

New rhythm control therapies are also emerging from clinical trials. In 2009, dronedarone, a novel antiarrhythmic, was approved by the Food and Drug Administration for the reduction of cardiovascular hospitalization in the treatment of AF.[8] Catheter-based strategies for the treatment of AF continue to evolve and are being deployed in wider patient populations, including patients with heart failure.[13,35] Beyond rhythm control, new research challenges the paradigm that strict rate control leads to improved outcomes in patients relegated to a rate-control strategy.[13,36] ORBIT-AF will also provide the opportunity to examine the impact of recent guideline changes regarding rate-control recommendations.[13] Although observational data, in general, cannot inform treatment decisions regarding efficacy, registry-based analyses can address several key components of quality care, including safety, comparative effectiveness, and equity.[37] The coordinated collection of patient characteristics, laboratory parameters, concomitant pharmacotherapies, quality-of-life data, and cardiovascular outcomes will enable timely assessment of these emerging therapies in clinical practice, outside the controlled environment of clinical trials.

Quality of life in AF

Atrial fibrillation is associated with significantly impaired quality of life.[38] Accordingly, many treatment decisions in AF are predicated upon improving both short- and long-term quality of life.[2] Prior studies of quality of life have been informative but have suffered from several limitations, including relatively small sample size, restricted follow-up, and the lack of validated AF-specific qualify-of-life instruments.[39] The ORBIT-AF PRO cohort will allow for large-scale examination of AF-

related quality of life in a heterogeneous group of patients over long-term follow-up using a rigorously derived and validated quality-of-life instrument.[19] Future analyses will attempt to determine if quality of life differs according to the type/pattern of AF, treatment strategies, or comorbidity. Additional analyses will address whether these differences in AF-related quality of life (if observed) are associated with differences in physician decisions related to rhythm control and anticoagulation strategies. Similarly, the anticoagulation-related quality-of-life assessment (ACTS) may also identify monitoring strategies,[40] existing therapies, and emerging antithrombotic therapies that engender patient satisfaction and improved adherence. Taken together, the PRO data will provide much needed information regarding quality of life among patients with AF. These data can also serve as a reference population for indirect comparisons of quality of life in experimental studies or comparative effectiveness research.

Limitations

As with any observational study, this registry will have limitations. Participation in the registry is voluntary. The patients and practices enrolled may not be entirely representative. Participation in this study, particularly as a part of a quality improvement effort, may alter practice patterns at a study site.[41] However, prior studies suggest that registry observations are comparable with those for general clinical practice.[42-44]

Conclusions

Although recent advances in the treatment of AF show promise for improved outcomes, no data exist to show how the results of these trials have influenced or will influence clinical practice. Prior studies have suggested that the application of evidence-based therapies remains suboptimal in patients with AF. The ORBIT-AF registry will examine current practice patterns, quality of care, and associated outcomes in the management of AF. These data will inform comparative effectiveness, comparative safety, and future research to improve quality of life and outcomes in patients with AF.

Disclosures

The ORBIT-AF registry is sponsored by Ortho-McNeil Janssen Scientific Affairs, LLC, Raritan, NJ.

References

1. Miyasaka Y, Barnes ME, Gersh BJ, et al. Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. Circulation 2006;114:119-25.
2. Fuster V, Ryden LE, Cannom DS, et al. ACC/AHA/ESC 2006 Guidelines for the Management of Patients With Atrial Fibrillation—Executive summary: a report of the ACC/AHA Task Force on Practice Guidelines and the ESC Committee for Practice Guidelines. J Am Coll Cardiol 2006;48:854-906.
3. Kilander L, Andren B, Nyman H, et al. Atrial fibrillation is an independent determinant of low cognitive function: A cross-sectional study in elderly men. Stroke 1998;29:1816-20.
4. Wang TJ, Larson MG, Levy D, et al. Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the Framingham Heart Study. Circulation 2003;107:2920-5.
5. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. Stroke 1991;22:983-8.
6. Lee WC, Lamas GA, Balu S, et al. Direct treatment cost of atrial fibrillation in the elderly American population: a Medicare perspective. J Med Econ 2008;11:281-98.
7. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med 2009;361:1139-51.
8. Hohnloser SH, Crijns HJ, van Eickels M, et al. Effect of dronedarone on cardiovascular events in atrial fibrillation. N Engl J Med 2009;360:668-78.
9. Mahaffey KW. Rivaroxaban once-daily oral direct factor Xa inhibition compared with vitamin K antagonism for prevention of stroke and embolism trial in atrial fibrillation (ROCKET AF). American Heart Association Scientific Sessions; November 15, 2010; Chicago.
10. Douglas PS, Brennan JM, Anstrom KJ, et al. Clinical effectiveness of coronary stents in elderly persons: results from 262,700 Medicare patients in the ACC–National Cardiovascular Data Registry. J Am Coll Cardiol 2009;53:1629-41.
11. Gheorghiade M, Filippatos G. Reassessing treatment of acute heart failure syndromes: the ADHERE registry. Eur Heart J 2005;7:B13-19.
12. Estes III NA, Halperin JL, Calkins H, et al. ACC/AHA/Physician Consortium 2008 clinical performance measures for adults with nonvalvular atrial fibrillation or atrial flutter: a report of the ACC/AHA Task Force on Performance Measures and the Physician Consortium for Performance Improvement. Circulation 2008;117:1101-20.
13. Wann LS, Curtis AB, January CT, et al. ACCF/AHA/HRS focused update on the management of patients with atrial fibrillation (updating the 2006 guideline): a report of the ACC/AHA Task Force on Practice Guidelines. Circulation 2011;123:104-23.
14. Chow SC, Chang M, Pong A. Statistical consideration of adaptive methods in clinical development. J Biopharm Stat 2005;15:575-91.
15. Calkins H, Brugada J, Packer DL, et al. HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for personnel, policy, procedures, and follow-up: a report of the Heart Rhythm Society (HRS) Task Force on Catheter and Surgical Ablation of Atrial Fibrillation. Heart Rhythm 2007;4:816-61.
16. The ROCKET AF Study Investigators. Rivaroxaban-once daily, oral, direct factor Xa inhibition compared with vitamin K antagonism for prevention of stroke and embolism trial in atrial fibrillation: rationale and design of the ROCKET AF study. Am Heart J 2010;159:340-7.
17. Connolly SJ, Eikelboom J, Joyner C, et al. Apixaban in patients with atrial fibrillation. N Engl J Med 2011;364:806-17.
18. Schulman S, Kearon C. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. J Thromb Haemost 2005;3:692-4.
19. Spertus J, Dorian P, Bubien R, et al. Development and validation of the Atrial Fibrillation Effect on QualiTy-of-Life (AFEQT) questionnaire in patients with atrial fibrillation. Circ Arrhythm Electrophysiol 2011;4:15-25.

20. Lip GY, Nieuwlaat R, Pisters R, et al. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor–based approach: the Euro Heart Survey on Atrial Fibrillation. Chest 2009;137:263-72.
21. Pisters R, Lane DA, Nieuwlaat R, et al. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. Chest 2010;138: 1093-100.
22. Go AS, Hylek EM, Chang Y, et al. Anticoagulation therapy for stroke prevention in atrial fibrillation: how well do randomized trials translate into clinical practice? JAMA 2003;290:2685-92.
23. Albers GW, Diener HC, Frison L, et al. Ximelagatran vs warfarin for stroke prevention in patients with nonvalvular atrial fibrillation: a randomized trial. JAMA 2005;293:690-8.
24. Hylek EM, Evans-Molina C, Shea C, et al. Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. Circulation 2007;115:2689-96.
25. Institute of Medicine. Initial national priorities for comparative effectiveness research. Washington, DC: The National Academies Press; 2009.
26. Kowey PR, Reiffel JA, Myerburg R, et al. Warfarin and aspirin use in atrial fibrillation among practicing cardiologist (from the AFFECTS registry). Am J Cardiol 2010;105:1130-4.
27. Fang MC, Go AS, Chang Y, et al. Warfarin discontinuation after starting warfarin for atrial fibrillation. Circ Cardiovasc Qual Outcomes 2010;3:624-31.
28. Steg PG. The Realise AF registry: an international, observational, cross-sectional survey evaluating atrial fibrillation management and the cardiovascular risk profile of atrial fibrillation patients. Stockholm: European Society of Cardiology Congress; 2010.
29. Le Heuzey JY, Breithardt G, Camm J, et al. The RECORD-AF study: design, baseline data, and profile of patients according to chosen treatment strategy for atrial fibrillation. Am J Cardiol 2010;105: 687-93.
30. Nieuwlaat R, Prins MH, Le Heuzey JY, et al. Prognosis, disease progression, and treatment of atrial fibrillation patients during 1 year: follow-up of the Euro Heart Survey on Atrial Fibrillation. Eur Heart J 2008;29:1181-9.
31. Nieuwlaat R, Capucci A, Camm AJ, et al. Atrial fibrillation management: a prospective survey in ESC member countries: the Euro Heart Survey on atrial fibrillation. Eur Heart J 2005;26: 2422-34.
32. Go AS, Hylek EM, Borowsky LH, et al. Warfarin use among ambulatory patients with nonvalvular atrial fibrillation: the AnTicoagulation and Risk Factors In Atrial Fibrillation (ATRIA) study. Ann Intern Med 1999;131:927-34.
33. Piccini JP, Lopes RD, Mahaffey KW. Oral factor Xa inhibitors for the prevention of stroke in atrial fibrillation. Curr Opin Cardiol 2010;25: 312-20.
34. Holmes DR, Reddy VY, Turi ZG, et al. Percutaneous closure of the left atrial appendage versus warfarin therapy for prevention of stroke in patients with atrial fibrillation: a randomised non-inferiority trial. Lancet 2009;374:534-42.
35. Khan MN, Jais P, Cummings J, et al. Pulmonary-vein isolation for atrial fibrillation in patients with heart failure. N Engl J Med 2008; 359:1778-85.
36. Van Gelder IC, Groenveld HF, Crijns HJ, et al. Lenient versus strict rate control in patients with atrial fibrillation. N Engl J Med 2010;362: 1363-73.
37. Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: The National Academies Press; 2001.
38. Dorian P, Jung W, Newman D, et al. The impairment of health-related quality of life in patients with intermittent atrial fibrillation: implications for the assessment of investigational therapy. J Am Coll Cardiol 2000;36:1303-9.
39. Thrall G, Lane D, Carroll D, et al. Quality of life in patients with atrial fibrillation: a systematic review. Am J Med 2006;119:e441-8.
40. Matchar DB, Jacobson A, Dolor R, et al. Effect of home testing of international normalized ratio on clinical events. N Engl J Med 2010; 363:1608-20.
41. Muller MP, Detsky AS. Public reporting of hospital hand hygiene compliance—helpful or harmful? JAMA 2010;304:1116-7.
42. Curtis L, Hernandez A, Greiner M, et al. Validity of a national heart failure quality of care registry: comparison of Medicare patients in OPTIMIZE-HF versus non–OPTIMIZE-HF hospitals. Circulation 2007; 115:e595.
43. Heidenreich PA, Fonarow GC. Are registry hospitals different? A comparison of patients admitted to hospitals of a commercial heart failure registry with those from national and community cohorts. Am Heart J 2006;152:935-9.
44. Granger CB, Gersh BJ. Clinical trials and registries in cardiovascular disease: competitive or complementary? Eur Heart J 2010;31: 520-1.

Appendix A. Event definitions

Major adverse cardiac events will be defined as the occurrence of cardiovascular death, myocardial infarction, or stroke or non-CNS (systemic) embolism. Cardiovascular death will include (but not be limited to) deaths due to heart failure, sudden cardiac death, stroke, arrhythmia, conduction disorders, vascular events, and non-CNS emboli.

Stroke is defined as a new, sudden, focal neurologic deficit, which results from a presumed cardiovascular cause that is not reversible within 24 hours and is not found to be due to a primary brain parenchymal abnormality (eg, tumor, mass-effect, primary CNS infection). Strokes will be classified as hemorrhagic or ischemic by the site investigator. Events in which symptoms do not persist for 24 hours (and without imaging evidence of infarction) will be defined as a TIA.

Myocardial infarction will be defined as clinical signs and symptoms consistent with myocardial infarction, accompanied by cardiac biomarker elevation (eg, creatine kinase–MB or troponin above the upper limit of normal).

Sudden cardiac death will be adjudicated according to the modified Hinkle-Thaler criteria, as used in several landmark cardiovascular trials. Patients who are well and (1) have a witnessed sudden collapse or (2) are found dead but were known to be alive and well in the previous 24 hours (eg, no signs or symptoms of cardiorespiratory distress) will be considered as sudden cardiac deaths.

Major bleeding will be adjudicated according to the International Society of Thrombosis and Haemostasis criteria. *Major bleeding* will be defined as clinically overt bleeding that is associated with any of the following: (1) a fall in hemoglobin level of $\geq$2 g/dL; (2) a transfusion of $\geq$2 U of packed red blood cells or whole blood; (3) bleeding in a critical site: intracranial, intraspinal, intraocular, pericardial, intra-articular, intramuscular with compartment syndrome, retroperitoneal; or (4) a fatal outcome. The 4 qualifying criteria will also be captured to allow for further classification of bleeding events.

Appendix B. Executive and steering committee members

An Executive Committee will govern the operations of ORBIT-AF. The Steering Committee will comprise the members of the executive committee as well as cardiologists, electrophysiologists, internists, and other hospital personnel from all regions of the United States. Responsibilities of the steering committee include program development, regional support and recruitment of principal investigators, participation in regional educational meetings, generation of publication topics/ideas, and participation in regional quality improvement initiatives.

Executive committee	Steering committee
Eric D. Peterson, MD, MPH, Chair	Larry A. Allen, MD, MHS
Jack Ansell, MD	Paul S. Chan, MD
Gregg C. Fonarow, MD	Michael D. Ezekowitz, MB, ChB
Bernard J. Gersh, MB, ChB, DPhil	James Freeman, MD
Alan S. Go, MD	Gerald Naccarelli, MD
Elaine Hylek, MD	James A. Reiffel, MD
Peter R. Kowey, MD	Steven Rothman, MD
Kenneth W. Mahaffey, MD	Daniel E. Singer, MD
Roger M. Mills, MD	
Jonathan P. Piccini, MD, MHS	
Laine Elliott Thomas, PhD	

BMC Health Services Research

RESEARCH ARTICLE Open Access

The quality of care for adults with epilepsy: an initial glimpse using the QUIET measure

Mary Jo Pugh[1,2*], Dan R Berlowitz[3,4], Jaya K Rao[5], Gabriel Shapiro[4,6], Ruzan Avetisyan[4], Amresh Hanchate[7], Kelli Jarrett[4], Jeffrey Tabares[1], Lewis E Kazis[3,8]

Abstract

Background: We examined the quality of adult epilepsy care using the *Quality Indicators in Epilepsy Treatment (QUIET)* measure, and variations in quality based on the source of epilepsy care.

Methods: We identified 311 individuals with epilepsy diagnosis between 2004 and 2007 in a tertiary medical center in New England. We abstracted medical charts to identify the extent to which participants received quality indicator (QI) concordant care for individual QI's and the proportion of recommended care processes completed for different aspects of epilepsy care over a two year period. Finally, we compared the proportion of recommended care processes completed for those receiving care only in primary care, neurology clinics, or care shared between primary care and neurology providers.

Results: The mean proportion of concordant care by indicator was 55.6 (standard deviation = 31.5). Of the 1985 possible care processes, 877 (44.2%) were performed; care specific to women had the lowest concordance (37% vs. 42% [first seizure evaluation], 44% [initial epilepsy treatment], 45% [chronic care]). Individuals receiving shared care had more aspects of QI concordant care performed than did those receiving neurology care for initial treatment (53% vs. 43%; $X^2 = 9.0$; $p = 0.01$) and chronic epilepsy care (55% vs. 42%; $X^2 = 30.2$; $p < 0.001$).

Conclusions: Similar to most other chronic diseases, less than half of recommended care processes were performed. Further investigation is needed to understand whether a shared-care model enhances quality of care, and if so, how it leads to improvements in quality.

Background

While existing quality indicators have focused on a number of highly prevalent chronic conditions (e.g., diabetes, hypertension) they do not address the quality of care for less prevalent, but serious conditions, such as epilepsy. Epilepsy care presents complexity in the sense that providers must balance seizure control, adverse drug effects, and complicated issues associated with epilepsy itself (e.g. mood disorders [1-3]) while also being mindful of consequences related to long-term treatment with antiepileptic drugs (e.g. bone health [4-6]). Thus, it is important to begin examining the quality of care provided to patients with epilepsy using quality measures and identifying gaps in quality of care. The United Kingdom has begun this process [7] due to the availability of not only clinical guidelines for care for patients with epilepsy[8,9], but also quality indicators from the Quality and Outcomes Framework [7]. While no comprehensive national guidelines for care of patients with epilepsy exist in the United States, the development of the QUality Indicators for Epilepsy Treatment in adults (QUIET) allows us to begin to examine the quality of epilepsy care in the United States.

The purpose of this study is to describe the quality of care received by adults with epilepsy in a major medical center in a Northeastern US city using the QUIET indicators-quality indicators developed as part of a larger study funded by the Centers for Disease Control and Prevention (CDC; Additional file 1)-and to assess the quality of epilepsy care in primary care and general neurology settings. Similar to other countries, in the US a substantial number of patients continue to receive their epilepsy care solely within the context of primary care

* Correspondence: pughm@uthscsa.edu
[1]South Texas Veterans Health Care System, VERDICT REAP, (7400 Merton Minter), San Antonio, TX, (78229), USA
Full list of author information is available at the end of the article

(55% in one study) [10]. Studies of quality of care for other chronic diseases have found better quality of care among patients receiving care from medical sub-specialists or within a shared care context [11-13]. Thus, we examine the extent to which variations exist in quality of care among patients who received epilepsy care only within primary care, only within neurology subspecialty care, and within both neurology and primary care (shared). Based on findings from previous studies, we hypothesize that patients with epilepsy are more likely to receive high quality care when they receive specialty care exclusively or have epilepsy care shared by both primary care and neurology specialty care [14-16].

Methods

Data

Data from the electronic medical record of a single medical center in the northeastern United States were used in this study to identify patients with epilepsy and assess the extent to which recommended processes of care were performed. The electronic medical record includes templates for certain aspects of care such as vital signs, medications, and lab tests. However, as progress notes are not disease specific or used for examining quality of care at the institution, they are primarily free text, which allows substantial variation in documentation of the care provided. Data were acquired from the demographic information, diagnosis codes, patient problem list, pharmacy, laboratory, inpatient, and outpatient components of the medical record. These data were entered into a specially designed chart abstraction form, and entered into a spreadsheet and exported to SPSS (Version 17.0, Chicago, IL) for subsequent analysis. This study was approved by the Institutional Review Board of the Medical Center.

Sample

Eligible participants received care at the medical center, were at least 18 years of age, able to speak English and, due to an additional component of the larger study, were able to complete a telephone interview. Probable epilepsy patients were identified by searching the electronic medical record system for all patients who had at least one ICD-9 code of 345.x or 780.39, or mention of the terms “epilepsy” or “seizure disorder” in the medical record problem list between 2004 and 2007. In order to assure there would be adequate data to assess the quality of epilepsy care we further required that individuals have two or more visits to the primary care or neurology clinics, or one visit to the primary care or neurology clinic and at least one hospitalization [17] between 2004 and 2007.

Individuals identified as eligible for the study based on diagnosis of epilepsy and who were confirmed to have epilepsy by their primary care provider or neurologist and received a letter from their physician informing them about the study and inviting their participation. Chart abstractions examining the quality of epilepsy care were completed by trained clinical chart abstractors only for those participants who contacted the study team after receiving the letter and provided written consent to participate in the study.

Quality Indicators

QUIET quality indicators were developed using the RAND appropriateness method-a process that integrates a systemic literature review with an expert consensus process to identify items that are considered appropriate and feasible measures of quality healthcare. Since existing evidence based guidelines from international sources existed, we adapted some items from those existing guidelines (e.g,, National Institute for Clinical Excellence [NICE] and Scottish Intercollegiate Guidelines Network [SIGN]), for use in the US healthcare setting [8,9]. Other items were developed using systematic literature review. A panel of 10 epilepsy experts and 1 primary care provider completed three rounds of ratings to identify items for which there was consensus regarding appropriateness, feasibility and necessity. Additional file 1 shows the QI’s that were rated appropriate and feasible indicators of epilepsy care quality, and further identifies QI’s that were adapted from the NICE or SIGN guidelines, and items which are parallel to measures included in the Quality and Outcomes Framework measure used in the United Kingdom [18]. Additional detail about this process is provided elsewhere [19].

Because of the large number of individual evidence-based QI’s during the development of the quality measures, we categorized the 22 QI’s into four aspects of epilepsy care: First Seizure Assessment (3 QI’s), Initial Epilepsy Treatment (7 QI’s), Chronic Epilepsy Care (9 QI’s), and Aspects of Care Unique to Women (3 QI’s). Additional file 1 shows specific QI’s for each aspect of care.

Operational definitions of quality indicators

Prior to chart abstraction, the research team consisting of physicians, nurses and health services researchers with expertise in conducting chart abstractions created operational definitions that identified specific data elements that would be used to score each QI. These definitions were then used to determine if applicable conditions were present for measuring each QI (i.e. the IF portion of the quality indicator), and then whether the process of care defined by the quality indicator was provided (i.e. the THEN portion). For example: IF the patient is diagnosed with a seizure disorder/epilepsy and started on therapy (denominator) THEN the patient

should be treated with monotherapy (numerator). Thus, there was variation in the denominator by QI. While some QI's focused on a single visit (at the time of the first seizure), most QI's examined the construct in question over time. For instance, a number of QI's recommended that a certain type of care be provided on a yearly basis (e.g. yearly depression screening). For these, we assessed visits within the designated time frame. Thus, our analysis is at the patient level for each QI, and the total number of individuals who met criteria for each QI was unique.

A specially designed chart abstraction instrument was developed, and data regarding processes of care provided was collected and entered using Microsoft Access. Research assistants with clinical background were specifically trained to perform chart reviews (Additional file 2). The final chart abstraction instrument was subjected to multiple layers of review by the research team. Initially, 25 charts were reviewed by two raters to assess concordance. Comparison of ratings for the quality indicators in those charts found that raters agreed 76% of the time. After finalization of the instrument, a second review of 25 charts conducted by independent raters found that there was agreement on 86% of ratings.

Initial assessment of concordance between QI's and care provided used an all-or- nothing approach: if all aspects of care were performed, the patient received a positive score for that indicator unless the indicator specifically stated that one of several aspects of care would fulfill the requirement (e.g. QI15; see Additional file 1). However, we also examined the proportion of recommended processes of care completed for each of the four aspects of epilepsy care described above.

Patient Characteristics

Demographic characteristics including age, sex, race/ethnicity, marital status, and education were abstracted from the electronic medical record. Epilepsy history (new-onset vs. chronic care), type of seizure (if documented), and medication information was ascertained based on longitudinal review of medical records and intake history for patients new to the medical center. Finally, the presence of continued seizures was identified by documentation of seizures in the medical record during the study period.

Setting of Epilepsy Care

Epilepsy care was identified by review of outpatient progress notes that described epilepsy and epilepsy care. For each epilepsy-focused medical encounter (encounter in which progress notes mentioned epilepsy) within the healthcare system, the type of provider was defined as being a primary care provider (general internist, family practitioner, or nurse practitioner), neurologist, or other specialty care. Individuals who received care only within primary care or neurology settings were classified as such. Those receiving epilepsy care in both settings were classified as receiving shared care.

Analysis

We first describe the variation in concordance among individual QI's (i.e., the concordance between the recommended and actual care received), followed by examination of the proportion of patients who received the care outlined for each aspect of epilepsy care (First Seizure, Initial Epilepsy Treatment, Chronic Care, Aspects of Care Unique to Women; Additional file 1 shows indicators included in each aspect of care). This analysis was conducted at the patient level, but included all recommended aspects of care for which each patient was eligible based on the individual indicators in each aspect of epilepsy care. This process is described in more detail below. Finally, we examine receipt of care by the setting of epilepsy care received using the chi square statistic. Haberman's adjusted residual statistic was used to identify significant cells within the chi-square analysis [20]. SPSS® version 17 (Chicago, IL, USA) was used to conduct data analysis.

Results

The sample for this study consisted of 311 individuals. Table 1 shows demographics for the sample overall and by source of care. Overall, approximately 58% were women, 62% were between 18 and 49 years of age; the sample was racially diverse with similar numbers of whites and African Americans. Over half had some college or a college degree. With regard to epilepsy, approximately 21% had new-onset epilepsy (within the past two years) and the median number of antiepileptic drugs (AEDs) prescribed at the last visit for this sample was 1 (mean = 1.41, SD = .90). Forty seven percent continued to have seizures during the study period, and 66% had no change in AED during the course of the study. Examination of demographic characteristics (Table 1) by source of care found that African Americans were more likely to be in shared care and primary care groups and less likely to be in the neurology group than expected (X^2 = 18.8, df = 2; $p < 0.01$). Individuals with continued seizures were less likely to be the primary care group and more likely to receive shared care than expected by chance (X^2 = 21.5, df = 2; $p < 0.01$). Individuals within the primary care group were less likely to have a medication change over the course of the study than expected by chance (X^2 = 9.6, df = 2; $p = 0.01$).

There was substantial variation in concordance between recommended care and actual care for individual QI's (Table 2). Excluding indicators with a total

Table 1 Characteristics of Study Participants

	All	Neurology Only N (%)	Shared care N (%)	Primary Care Only N (%)
Total	311	203 (65.27)	77 (24.76)	31 (9.97)
Sex				
Female	181 (58.20)	114 (56.16)	50 (64.94)	17 (54.84)
Male	130 (41.80)	89 (43.84)	27 (35.06)	14 (45.16)
Age				
18-49	194 (62.38)	131 (64.53)	46 (59.74)	17 (54.84)
50-64	80 (25.72)	46 (22.66)	24 (31.17)	10 (32.26)
65+	37 (11.90)	26 (12.81)	7 (9.09)	4 (12.90)
Race/Ethnicity				
White	147 (47.27)	112 (55.17)*	27 (35.06)*	8 (25.81)*
African American	133 (42.77)	70 (34.48)*	42 (54.55)*	21 (67.74)*
Other	31 (9.97)	21 (10.34)	8 (10.39)	2 (6.45)
Education				
Less than high school	50 (16.29)	28 (13.86)	13 (17.33)	9 (30.00)
High school graduate	91 (29.64)	63 (31.19)	23 (30.67)	5 (16.67)
Some college	93 (30.29)	57 (28.22)	23 (30.67)	13 (43.33)
College graduate	73 (23.78)	54 (26.73)	16 (21.33)	3 (10.00)
New-Onset Epilepsy	65 (20.90)	45 (21.17)	16 (20.78)	4 (12.90)
Number of Epilepsy Medications (Last clinic visit)				
0	32 (10.29)	19 (9.36)	7 (9.09)	6 (19.35)
1	166 (53.38)	109 (53.69)	41 (53.25)	16 (51.61)
2	76 (24.44)	51 (25.12)	19 (24.68)	6 (19.35)
3 or more	37 (11.90)	24 (11.83)	10 (12.99)	3 (9.68)
No Antiepileptic Drug Change	207 (65.56)	133 (65.52)	46 (59.74)	28 (90.32)
Continued seizures	147 (47.27)	90 (44.33)*	51 (66.23)	6 (19.35)*

Education is self-reported

*$p < 0.01$.

denominator of less than 25 (to ensure precision of analyses) concordance ranged from 2% for QI16 to 99% for QI's 15 and 18. The mean proportion of concordant care by indicator was 54.0 (standard deviation = 28.0); the median was 59.26.

Table 2 shows the proportion of patients who received QI concordant care for each QI overall and by the setting of epilepsy care. The only QI's where sufficient primary care patients were represented for the purposes of analysis were in chronic epilepsy care: QI 14, 15 16, 20, 21, and 23. For these aspects of chronic epilepsy care we found no statistically significant differences in the extent to which patients received referral for treatment after a positive depression screen (QI 21; $X^2 = 0.73$, df = 2; p = 0.7), or folate supplementation for women of childbearing age (QI 23: $X^2 = 3.6$, df = 2; p = 0.2) by the setting of epilepsy care. Haberman's Adjusted Residual analysis of statistically significant analyses indicated that patients who received primary care were less likely than expected to have documentation of approximate seizure count since their last visit (QI 14; $X^2 = 13.6$, df = 2; p = 0.001) or interventions performed in light of continued seizures documentation of patient education (QI 15: $X^2 = 47.6$, df = 2; $p < 0.001$). Individuals who received care in neurology and primary care settings were less likely to receive depression screening than those who received shared care (QI 20 $X^2 = 6.2$, df = 2; 26.21; $p < 0.001$).

Table 3 shows the proportion of QI concordant care among the four aspects of epilepsy care (first seizure evaluation, initial epilepsy treatment, chronic epilepsy care, and aspects of care unique to women) overall and by setting of epilepsy care. The number reported in each column represents the proportion of all possible care processes that were performed among the care processes outlined by the QI for those who met QI inclusion criteria. For evaluation of a first seizure, of the 65 individuals who received a first seizure evaluation there were 151 possible opportunities for care among individuals who met inclusion criteria regarding QI's for first seizure assessment (QI 1-3). Of those 151 opportunities for QI concordant care, 64 were completed (42.38%). Among the 65 individuals who met criteria for QI's examining initial treatment of epilepsy, there were 297

Table 2 Proportion of Patients Receiving QI Concordant Care by Setting of Care

Quality Indicator	All N = 311	Neurology only N = 203	Shared care N = 77	Primary Care Only N = 31
Evaluation of First Seizure				
QI 1.	22/65 = 33.85%	11/45 = 24.44%	10/16 = 62.5%	1/4 = 25%
QI 2.	25/65 = 38.46%	18/45 = 40%	7/16 = 43.75%	0/4 = 0%
QI 3.	17/21 = 80.95%	12/16 = 75%	3/3 = 100%	2/2 = 100%
Initial Treatment of Epilepsy				
QI 4.	29/65 = 44.62%	21/45 = 46.67%	8/16 = 50%	0/4 = 0%
QI 5.	10/65 = 15.38%	5/45 = 11.11%	4/16 = 25%	1/4 = 25%
QI 6.	34/46 = 73.91%	23/31 = 74.19%	10/11 = 90.91%	1/4 = 25%
QI 7.	21/23 = 91.3%	15/15 = 100%	6/6 = 100%	0/2 = 0%
QI 8.	17/64 = 26.56%	12/44 = 27.27%	5/16 = 31.25%	0/4 = 0%
QI 9.	NA	NA	NA	NA
QI 11.	20/31 = 64.52%	12/21 = 57.14%	7/9 = 77.78%	1/1 = 100%
Follow-up/Chronic Disease Care				
QI 14.	76/272 = 27.94%	53/176 = 30.11%	23/66 = 34.85%	0/30 = 0%
QI 15.	145/147 = 98.64%	90/90 = 100%	51/51 = 100%	4/6 = 66.67%
QI 16	6/311 = 1.93%	2/203 = 0.99%	4/77 = 5.19%	0/31 = 0%
QI 17.	16/27 = 59.26%	12/19 = 63.16%	4/8 = 50%	
QI 18.	69/70 = 98.57%	43/43 = 100%	26/27 = 96.3%	
QI 19.	82/132 = 62.12%	50/83 = 60.24%	30/46 = 65.22%	2/3 = 66.67%
QI 20.	142/311 = 45.66%	72/203 = 35.47%	53/77 = 68.83%	17/31 = 54.84%
QI 21.	88/125 = 70.4%	53/73 = 72.6%	22/34 = 64.71%	13/18 = 72.22%
QI 22.	11/14 = 78.57%	6/7 = 85.71%	5/7 = 71.43%	
Aspects of Care Specific to Women				
QI 23. I	38/111 = 34.23%	24/73 = 32.88%	13/29 = 44.83%	1/9 = 11.11%
QI 24.	1/5 = 20%	0/2 = 0%	1/3 = 33.33%	
QI 25.	8/12 = 66.67%	3/6 = 50%	5/5 = 100%	

QI number is based on the original QI's reported in Pugh MJ, Berlowitz DR, Montouris G *et al*. What constitutes high quality of care for adults with epilepsy? Neurology 2007; 69(21):2020-7. QI were not deemed appropriate and necessary (QI 10, 12, 13) are not presented in this paper.

opportunities for recommended care. QI concordant care was provided for 131 (44.11%) of those opportunities. Among the 311 individuals who met criteria for many of the chronic epilepsy care QI's, there were 1,409 opportunities for recommended care; 45.07% of the time QI concordant care was provided. Among the 111 women of child-bearing age included in the study there were 128 opportunities for recommended care; 36.72% of the time QI concordant care was provided. These data indicate that overall, less than half of all possible QI identified care processes (877/1985; 44.2%) were completed in this sample. The lowest concordance between recommended care and actual care was for aspects of care unique to women.

Further examination of quality within each aspect of epilepsy care suggests that there were also significant differences by the setting in which care was received. Our examination of differences between settings of care was restricted to neurology vs. shared care for First Seizure Assessment and Aspects of Care Unique to

Table 3 Proportion of all Possible Opportunities Taken for Quality Epilepsy Care by Setting of Epilepsy Care

Aspect of Epilepsy Care	N***	All	Neurology Only	Shared care	Primary Care Only
First seizure assessment[$]	63	64/151 = 42.38%	41/106 = 38.70%	20/35 = 57.14%	3/10 = 30.00%
Initial epilepsy treatment	63	131/297 = 44.11%	88/203 = 43.35%	40/75 = 53.33%	**3/19 = 15.79%**
Chronic epilepsy care	302	635/1409 = 45.07%	381/897 = 42.47%	**218/393 = 55.47%**	36/119 = 30.25%
Aspects of care unique to women[$]	108	47/128 = 36.72%	27/81 = 33.33%	19/37 = 51.35%	1/10 = 10.00%

[$] Insufficient N for Primary Care Only; unable to compare.

* $p < 0.05$.

** $p < .01$.

***N that could potentially qualify for inclusion in analysis if all inclusion criteria for a specific quality indicator are met.

The numerator in each column represents the number of possible care processes that were performed; the denominator represents the number of possible care processes. For specific numbers of patients, and specific indicators included in each aspect of epilepsy care please refer to Table 2.

Women due to low numbers of patients receiving Primary Care Only. For First Seizure Assessment, we found a trend approaching significance with individuals receiving shared care having more aspects of QI concordant care performed than did those receiving care in a neurology setting (Table 3; $X^2 = 3.7$; $p = 0.06$). For Initial Epilepsy Treatment concordance by setting of care varied from 16% for those receiving primary care only to 53% for those receiving shared care; individuals receiving primary care were significantly less likely to receive QI concordant care than expected ($X^2 = 9.0$; $p = 0.01$). Concordance for QI's included in Chronic Epilepsy Care ranged from 30% for those receiving primary care only to 55% for those receiving shared care. Individuals receiving shared care were more likely to receive QI concordant care than those receiving neurology or primary care ($X^2 = 30.2$; $p < 0.001$). For Aspects of Care Unique to Women, there was a trend approaching significance, with individuals receiving shared care being more likely to receive QI concordant care than those receiving neurology care ($X^2 = 3.5$; $p = 0.06$).

Discussion

This study used data from electronic medical records to assess the extent to which patients with epilepsy receive processes of epilepsy care identified as QI's designed for use in primary and general neurology care. We were able to reliably assess the extent to which recommended care was documented in patient records. However findings for several indicators where fewer than 5% of the sample received recommended aspects of care suggest additional evaluation of those indicators or data sources is needed.

Consistent with studies examining QI concordant care in other chronic diseases, less than half of all possible care processes were completed [21]. There was wide variation in the extent to which all recommended processes of care were provided, with Aspects of Care Unique to Women having the lowest rates of concordance.

Our data provide limited support for our hypothesis that individuals receiving shared care have better quality of care than individuals receiving only primary care or only neurology subspecialty care. The only QI's for which there were significant differences between patients receiving neurology care and those shared care was QI20 which suggests that persons with epilepsy should receive an annual depression assessment. Our data do not indicate who conducted the depression assessment. It is possible that the neurologist was more likely to attend to these items immediately for a patient that would be seen only once for referral, than the more complex patients with competing demands who were seen continuously in neurology settings [22]. The more complex patients who were seen exclusively in neurology settings likely had acute issues associated with their seizures that required immediate attention, leaving little or no time at the end of an office visit to address chronic disease management issues. Alternatively, these chronic disease management issues may have also been addressed by the neurologist but not documented due to the intensive documentation required to address more acute seizure care.

Further examination of the aggregate measures where there is more power to detect differences between individuals receiving shared vs. primary or neurology care only, our analysis revealed that there was a significant difference between shared care and neurology only care only for chronic disease management.

Shared care has been examined in a variety of contexts with mixed results. A number of studies have found no difference in quality for patients in shared care compared to those in specialty care exclusively [23,24]. Rosendal and colleagues found that individuals with hip fracture and receiving shared care had more home care after discharge, and lower scores on the short version of the Sickness Impact Profile indicating improved recovery [25]. Pugh and colleagues found that patients with diabetes were more likely to receive guideline concordant diabetes medications if they received care from both primary and specialty care compared to primary care or specialty care exclusively [26].

Our analysis suggests that, for epilepsy, QI's associated with types of care that are more technical in nature (e.g., QI 2, 7, 15, 18) tended to have higher rates of concordance, while those associated with discussion, patient education, and chronic disease management (e.g., QI 5, 8, 16) tended to have lower rates of concordance. It is possible that these aspects of patient care that are more interpersonally oriented were actually performed, but not documented. However, documentation of these aspects of care is in and of itself an indicator of quality. Because several of these interpersonally oriented items are ones included in the UK's Quality and Outcomes framework, there is broad international consensus that they are critical components of the quality of epilepsy care (e.g., particularly documenting seizure frequency and reviewing medication management). Accordingly the QUIET indicators could be used to develop electronic templates for use in documenting care for patients with epilepsy in organizations with electronic medical records (or a paper template for settings with paper records). Such a template would facilitate documentation of these important processes of care when they occur, and guide providers who do not commonly provide care for patients with epilepsy.

Beyond our hypothesis, we found that chronic disease management issues specific to women tended to have lower rates of concordance than other aspects of care. Again, it is possible that women were taking over-

the-counter medications or vitamins with the recommended dose of folate and did not need a prescription. However, the drug-drug interaction between lamotrigine and oral contraceptives was not addressed in four of five eligible patients. Because the numbers of women meeting criteria was low for QI 24 women's issues should be highlighted as a concern not only for clinicians to address, but also as a concern requiring additional quality assessment.

Finally, our findings suggest that issues recently highlighted in the literature-bone health and mood disorders [27,28] -are being addressed for many patients with epilepsy. Approximately 62% of individuals on AEDs for two or more years received testing for Vitamin D or a DXA scan. With regard to mood disorders, 46% received recommended screening. Of those with evidence of anxiety or depression, 71% received treatment with medications or a referral to a mental health practitioner.

These findings must be interpreted in light of several limitations. First, this study was in a single medical center in which study participants appeared to have higher rates of uncontrolled seizures than estimates from the literature [29]. The complexity of these patients may be such that competing demands for physician time make addressing less acute issues more difficult, so these findings may not reflect the quality of care provided for less complicated patients. Attempts to adjust these results for disease severity, however, did not change results since there were no significant differences in QI concordant care for those with and without continued seizures. This limited sample, from a single geographic location provides a point of reference for future studies from other geographic regions and multi-site studies with power to detect geographic variations in care that may exist.

Related to this single-site study is the fact that the total number of patients (N = 311) was relatively small, particularly after stratifying patients into groups of neurology care, shared care and primary care. The primary care group was particularly small for all groups except chronic epilepsy care. Thus, findings for the primary care group must be interpreted with caution.

Next, quality of care was assessed primarily using the electronic medical record. As described above, it is possible that interpersonal aspects of care were addressed but not documented. A number of aspects of care that are more interpersonal in nature may be better assessed using patient reported survey measures. Future studies examining the quality of epilepsy care using a broad measure such as the QUIET indicators may benefit from an approach that utilizes not only medical chart abstraction, but a targeted survey that provides a second assessment of patient reported receipt of care which allows one to triangulate data to assess the care received by patients. Moreover, such a survey would benefit by adding quality of life measures, and other potential outcomes that may be linked to higher quality care such as employment, relationship status, and education.

We further found that scoring compound QI's (e.g., 1, 8 and 16) in an all or nothing fashion may lead to loss of useful information. Adaptation and/or disaggregation of specific aspects of these QI's may be needed in future research on the quality of epilepsy care. However, our assessment of broad areas of epilepsy care (Table 3) examined the proportion of possible care processes that were completed. Even in this analysis where credit was given for completing even one aspect of care for complex QI's, fewer than half of all possible processes of care were completed, suggesting room for improvement in all aspects of epilepsy care.

Finally, because these QI's were designed for use in primary or general neurology care settings, it is possible that some very important aspects of care may not be as clearly articulated. For instance we indicated numerous points where referral to a higher level of specialty care is recommended. Due to the structure of care in the US, we did not specifically identify a point in time where a patient should be referred to a tertiary epilepsy center for comprehensive evaluation. This limitation should be addressed by future versions of this measure.

Conclusion

This study provides a snapshot of the quality of care provided to adults with epilepsy in one metropolitan healthcare system. While results are only from a single hospital, these findings suggest that assessment of quality using most of the QUIET QI's is feasible, and that care for epilepsy can in large part be reliably measured using medical chart abstraction. This tool will allow identification of gaps in quality for epilepsy care within other healthcare systems, and eventually it can be used to improve the care provided for adults with epilepsy. However additional evaluation of the QUIET measure using patient surveys and development of QI's for use in specialty care is also needed.

Additional material

Additional file 1: Quality in epilepsy treatment in adults (QUIET) indicators. This file includes the specific quality indicators included in the QUIET measure.

Additional file 2: CDC epilepsy quality of care study: patients' medical record review/abstraction form. This file includes the chart abstraction form used to collect data on quality of care from patients' electronic medical record.

Acknowledgements

Funding for this study was provided by the Centers for Disease Control and Prevention (PI: Dr. Lewis Kazis). We thank David Thurman, MD, MPH for his thoughtful review of this manuscript. This material is the result of work

supported with resources and the use of facilities at the South Texas Veterans Health Care System and the Center for Health Quality, Outcomes and Economic Research (CHQOER), Edith Nourse Rogers VA Medical Center Hospital. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Centers for Disease Control and Prevention or the Department of Veterans Affairs or the United States government.

Author details
[1]South Texas Veterans Health Care System, VERDICT REAP, (7400 Merton Minter), San Antonio, TX, (78229), USA. [2]University of Texas Health Science Center at San Antonio, Department of Epidemiology and Biostatistics, (7703 Floyd Curl Drive), San Antonio, TX, (78229) USA. [3]Center for Health Quality, Outcomes and Economic Research (CHQOER), Veterans Administration Medical Center, (200 Springs Road), Bedford, MA, (01730), USA. [4]Boston University School of Public Health, Health Policy and Management, (715 Albany St, Talbot Building), Boston, MA, (02118), USA. [5]University of North Carolina, Eshelman School of Pharmacy, (Kerr Hall, Room 2202) Chapel Hill, NC, (27599), USA. [6]Hôpital Sainte-Justine, Université de Montréal, (3175, Chemin de la Côte-Sainte-Catherine), Montreal, Quebec, (H3T 1C5), Canada. [7]Boston Veterans Administration Medical Center, Center for Organization, Leadership, and Management Research, (150 South Huntington Ave.), Jamaica Plain, MA, (02130), USA. [8]Center for the Assessment of Pharmaceutical Practices (CAPP) and Pharmaceutical Assessment Management and Policy (PAMP), Department of Health Policy and Management, Boston University School of Public Health, (715 Albany St, Talbot Building), Boston, MA, (02118), USA.

Authors' contributions
MJP contributed to acquisition of funding, concept and design of the study, interpretation of data, and preparation of manuscript. DRB contributed to acquisition of funding, concept and design of the study, analysis and interpretation of data, and preparation of manuscript. GS MJP contributed to data acquisition, analysis of data, and preparation of manuscript. GDM MJP contributed to interpretation of data, and preparation of manuscript. RA contributed to data acquisition and preparation of manuscript. AH contributed to analysis and interpretation of data and preparation of manuscript. KJ MJP contributed to data acquisition and preparation of manuscript. JT contributed to analysis of data and preparation of manuscript. LEK contributed to acquisition of funding, concept and design of the study, interpretation of data, and preparation of manuscript. All authors read and approved the final manuscript.

Competing interests
Dr. Pugh reports no disclosures
Dr. Berlowitz reports no disclosures
Mr. Shapiro reports no disclosures
Dr. Montouris reports no disclosures
Dr. Avetisyan reports no disclosures
Dr. Hanchate reports no disclosures
Ms. Jarrett reports no disclosures
Mr. Tabares reports no disclosures
Dr. Kazis reports receiving research funding for other projects from Eli Lilly and Company, Bristol Meyers Squibb, Sanofi-Aventis, Genzyme, Boehringer-Ingelheim, Amgen, and Janssen Pharmaceuticals

Received: 30 July 2010 Accepted: 3 January 2011
Published: 3 January 2011

References

1. Jagadheesan K, Garg AK, Nizamies SH: **Risk factors and outcome of mood disorders in epilepsy: a case-control study.** *Seizure* 2003, **12**:121-125.
2. Gaitatzis A, Trimble MR, Sander JW: **The psychiatric comorbidity of epilepsy.** *Acta Neurol Scand* 2004, **110**:207-220.
3. Cramer JA: **Mood disorders are linked to health-related quality of life in epilepsy.** *Epilepsy Behav* 2002, **3**:491-492.
4. Pack AM: **The association between antiepileptic drugs and bone disease.** *Epilepsy Curr* 2003, **3**:91-95.
5. Pack AM, Morrell MJ, Marcus R, Holloway L, Flaster E, Done S, Randall A, Seale C, Shane E: **Bone mass and turnover in women with epilepsy on antiepileptic drug monotherapy.** *Ann Neurol* 2005, **57**:252-257.
6. Cohen A, Lancman M, Mogul H, Marks S, Smith K: **Strategies to protect bone mass in the older patient with epilepsy.** *Geriatrics* 1997, **52**:70-81.
7. Shohet C, Yelloly J, Bingham P, Lyratzopoulos G: **The association between the quality of epilepsy management in primary care, general practice population deprivation status and epilepsy-related emergency hospitalisations.** *Seizure* 2007, **16**:351-355.
8. **National Collaborating Center for Primary Care.** [http://www.guideline.gov/content.aspx?id=5963].
9. **Diagnosis and management of epilepsy in adults.** [http://www.sign.ac.uk/guidelines/fulltext/70/index.html].
10. Pugh MJ, Cramer J, Knoefel J, Charbonneau A, Mandell A, Kazis L, Berlowitz D: **Potentially inappropriate antiepileptic drugs for elderly patients with epilepsy.** *J Am GerSoc* 2004, **52**:417-422.
11. Ayanian JZ, Guadognoli E, McNeil BJ, Cleary PD: **Treatment and outcomes of acute myocardial infarction among patients of cardiologists and generalist physicians.** *Arch Int Med* 1997, **157**:2570-2576.
12. Rosenblatt RA, Baldwin LM, Chan L, Fordyce MA, Hirsch IB, Palmer JP, Wright GE, Hart LG: **Improving the quality of outpatient care for older patients with diabetes: lessons from a comparison of rural and urban communities.** *J Fam Pract* 2001, **50**:676-680.
13. Cheng EM, Swarztrauber K, Siderowf AD, Eisa MS, Lee M, Vassar S, Jacob E, Vickrey BG: **Association of specialist involvement and quality of care for Parkinson's disease.** *Mov Disord* 2007, **22**:515-522.
14. Ayanian JZ, Hauptman PJ, Guadognoli E, Antman EM, Pashos CL, McNeil BJ: **Knowledge and practice of generalist and specialist physicians regarding drug therapy for acute myocardial infarction.** *New Engl J Med* 1994, **331**:1136-1142.
15. Chin MH, Zhang JX, Merrell K: **Specialty differences in the care of older patients with diabetes.** *Med Care* 2000, **38**:131-140.
16. Chin MH, Friedmann PD, Cassell CK, Lang RM: **Differences in generalist and specialist physicians' knowledge and use of angiotensin-converting enzyme inhibitors for congestive heart failure.** *J Gen Intern Med* 1997, **12**:523-530.
17. Hebert PL, Geiss LS, Tierney EF, Engelgau MM, Yawn BP, McBean AM: **Identifying persons with diabetes using medicare claims data.** *Am J Med Qual* 1999, **14**:270-277.
18. **Developing clinical and health improvement indicators for the Quality and Outcomes Framework (QOF) Interim process guide.** [http://www.nice.org.uk/media/742/32/QOFProcessGuide.pdf].
19. Pugh MJ, Berlowitz DR, Montouris G, Bokhour B, Cramer JA, Bohm VA, Bollinger MJ, Helmers SL, Ettinger AB, Meador KJ, Fountain NB, Boggs JG, Tatum WO, Knoefel J, Harden C, Mattson RH, Kazis LE: **What constitutes high quality of care for adults with epilepsy?** *Neurology* 2007, **69**:2020-2027.
20. Haberman SJ: *Analysis of Qualitative Data* Burlington: Academic Press; 1979.
21. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA: **The quality of health care delivered to adults in the United States.** *N Eng J Med* 2003, **348**:2635-2645.
22. Parchman ML, Pugh JA, Romero RL, Bowers KW: **Competing demands or clinical inertia: the case of elevated glycosylated hemoglobin.** *Ann Fam Med* 2007, **5**:196-201.
23. Page J, Weber R, Somaini B, Nostlinger C, Donath K, Jaccard R: **Quality of generalist vs. specialty care for people with HIV on antiretroviral treatment: a prospective cohort study.** *HIV Med* 2003, **4**:276-86.
24. Cheung NW, Yue DK, Kotowicz MA, Jones PA, Flack JR: **A comparison of diabetes clinics with different emphasis on routine care, complications assessment and shared care.** *Diabet Med* 2008, **25**:974-978.
25. Rosendal H, van Beekum WT, Nijhof P, de Witte LP, Schrijvers AJ: **Can shared care deliver better outcomes for patients undergoing total hip replacement?** *Int J Integr Care* 2000, **1**:1-8.
26. Pugh MJ, Anderson J, Pogach LM, Berlowitz DR: **Differential adoption of pharmacotherapy recommendations for type 2 diabetes by generalists and specialists.** *Med Care Res Rev* 2003, **60**:178-200.
27. Gidal BE, Sheth RD: **Epilepsy and bone health: is there cause for concern?** *Epilepsy Behav* 2004, **5**:1-2.

28. Gilliam FG: **Diagnosis and treatment of mood disorders in persons with epilepsy.** *Curr Opin Neurol* 2005, **18**:129-133.
29. Brodie MJ, Kwan P: **Staged approach to epilepsy management.** *Neurology* 2002, **58**:S2-S8.

Pre-publication history
The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1472-6963/11/1/prepub

doi:10.1186/1472-6963-11-1
Cite this article as: Pugh *et al.*: The quality of care for adults with epilepsy: an initial glimpse using the QUIET measure. *BMC Health Services Research* 2011 **11**:1.

Trends in Use of High-Risk Medications for Older Veterans: 2004 to 2006

Mary Jo V. Pugh, PhD,[ab] Joseph T. Hanlon, PharmD, MS,[cd] Chen-Pin Wang, PhD,[ab] Todd Semla, PharmD,[efg] Muriel Burk, PharmD,[efg] Megan E. Amuan, MPH,[h] Ashlei Lowery, MD,[i] Chester B. Good, MD,[dj] and Dan R. Berlowitz, MD, MPH[gk]

OBJECTIVES: To examine the change in use of high-risk medications for the elderly (HRME), as defined by the National Committee on Quality Assurance's Healthcare Effectiveness Data and Information Set (HEDIS) quality measure (HEDIS HRME), by older outpatient veterans over a 3-year period and to identify risk factors for HEDIS HRME exposure overall and for the most commonly used drug classes.

DESIGN: Longitudinal retrospective database analysis.

SETTING: Outpatient clinics within the Department of Veterans Affairs (VA).

PARTICIPANTS: Veterans aged 65 by October 1, 2003, and who received VA care at least once each year until September 30, 2006.

MEASUREMENTS: Rates of use of HEDIS HRME overall and according to specific drug classes each year from fiscal year 2004 (FY04) to FY06.

RESULTS: In a cohort of 1,567,467, high-risk medication exposure fell from 13.1% to 12.3% between FY04 and FY06 ($P<.001$). High-risk antihistamines (e.g., diphenhydramine), opioid analgesics (e.g., propoxyphene), skeletal muscle relaxants (e.g., cyclobenzaprine), psychotropics (e.g., long half-life benzodiazepines), endocrine (e.g., estrogen), and cardiac medications (e.g., short-acting nifedipine) had modest but statistically significant ($P<.001$) reductions (range -3.8% to -16.0%); nitrofurantoin demonstrated a statistically significant increase ($+36.5\%$; $P<.001$). Overall HEDIS HRME exposure was more likely for men, Hispanics, those receiving more medications, those with psychiatric comorbidity, and those without prior geriatric care. Exposure was lower for individuals exempt from copayment. Similar associations were seen between ethnicity, polypharmacy, psychiatric comorbidity, access-to-care factors, and use of individual HEDIS HRME classes.

CONCLUSION: HEDIS HRME drug exposure decreased slightly in an integrated healthcare system. Risk factors for exposure were not consistent across drug groups. Future studies should examine whether interventions to further reduce HEDIS HRME use improve health outcomes. **J Am Geriatr Soc 59:1891–1898, 2011.**

Key words: inappropriate prescribing; quality of care; aged; pharmacoepidemiology

Dr. Mark Beers first developed explicit criteria for potentially inappropriate prescribing in the elderly (PIPE), defined as medications rated by an expert panel as those whose risks outweigh their potential benefit in older adults, in 1991. They were subsequently updated in 1997 and 2003. Numerous studies have shown the rates of PIPE based on the Beers criteria over the past decade to be between 20% and 30% in older outpatients.[1–8] Several recent studies have identified small but significant reductions in PIPE as defined according to different versions of the Beers criteria.[9,10] For example, one study found a 3.9% reduction in PIPE (2003 Beers Criteria) between 1996 and 2005 in the United Kingdom despite a trend for an increase in the number of prescriptions overall during this time.[11]

In 2006, the National Committee on Quality Assurance (NCQA) developed a Healthcare Effectiveness Data

From the [a]South Texas Veterans Health Care System, Veterans Evidence-based Research and Implementation Center, San Antonio, Texas; [b]Department of Epidemiology and Biostatistics and [i]Division of Geriatrics, Gerontology and Palliative Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas; [c]Division of Geriatric Medicine, Departments of Pharmacy and Therapeutics and Epidemiology and [j]Division of General Internal Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania; [d]Geriatric Research Education and Clinical Center and Center for Health Equity Research and Promotion, Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, Pennsylvania; [e]Pharmacy Benefits Management Services, Department of Veterans Affairs, Hines, Illinois; Departments of [f]Medicine and [g]Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; [h]Center for Health Quality, Outcomes, and Economic Research, Department of Veterans Affairs, Bedford, Massachusetts; and [k]School of Public Health, Boston University, Boston, Massachusetts.

Address correspondence to Mary Jo Pugh, VERDICT REAP, South Texas Veterans Health Care System (11C6), 7400 Merton Minter, San Antonio, TX 78023. E-mail: maryjo.pugh2@va.gov or pughm@uthscsa.edu

DOI: 10.1111/j.1532-5415.2011.03559.x

0002-8614/11/$15.00

and Information Set (HEDIS) quality measure to examine use of high-risk medications in the elderly (HRME) developed by an expert panel and based in part on the most-recent version of the Beers criteria.[12,13] The HEDIS HRME measure included some, but not all, of the drugs included in the Beers criteria, retaining those for which there was consensus that they should be avoided and that outcomes were considered to be of high severity. This more-refined measure is now used to benchmark the quality of medication management in older adults enrolled in Medicare and other managed care plans and thus is of interest to the Department of Veterans Affairs, Veterans Health Administration (VA).

Using VA data from October 1, 1999 to September 30, 2000, it was previously reported that the overall 1-year prevalence of HEDIS HRME exposure in older veteran outpatients was 19.6%.[14] To the best of the knowledge of the authors of the current study, the only published studies examining longitudinal use of HEDIS HRME as defined by this HEDIS quality measure have been reports by the NCQA.[15]

The objective of the current study was to examine the change in HEDIS HRME exposure in older veteran outpatients between 2004 and 2006 and to examine risk factors for HEDIS HRME exposure. Given that the foundation for the HEDIS HRME measure existed for longer than a decade before the refinement of the measure itself, it was hypothesized that reductions in the rates of HEDIS HRME drug use overall and within specific drug classes should have been evident by the mid-2000s. Moreover, it was hypothesized that risk factors for HEDIS HRME would be similar to those reported in previous studies using the Beers criteria and that they would be consistent across the major drug groups that constitute the HEDIS HRME measure.

METHODS

Study Design, Setting, and Sample

A longitudinal retrospective data analysis study was conducted using data from all VA outpatient clinics. The sample consisted of veterans aged 65 by October 1, 2003 (beginning fiscal year 2004 (FY04)) who received VA care at least once each year between FY03 and FY06. To examine change in a consistent sample over time, individuals who received sporadic VA care or who died during this period were not included in the analysis. Institutional review boards at three sites (University of Texas Health Science Center at San Antonio, Hines VA, and Bedford VA) approved this study.

Data Sources

National VA inpatient, outpatient, and pharmacy data from FY03 (October 1, 2002–September 30, 2003) through FY06 (October 1, 2005–September 30, 2006) were obtained for individuals aged 65 and older at the beginning of FY04. A merged database was created using information from the VA National Patient Care Database records and all outpatient pharmacy prescription data from the VA Pharmacy Benefits Management database. Records were merged using an encrypted identifier that is consistent for each person across VA data sets.

Measures

Exposure: HRME

Use of any of the HEDIS HRME drugs was identified using VA pharmacy data each year between FY04 and FY06. A measure of exposure was then created for overarching groups of drugs based on the VA Medication class system (Table 1) (http://www.pbm.va.gov/natform/vaclass.xls).

Independent Variables

Patient demographic characteristics (age, sex, race and ethnicity) were identified using data fields from VA administrative databases between FY03 and FY06. With the exception of race, these demographic characteristics are well documented and complete in the medical record. Because the process of recording race changed in 2002, race data are more likely to be missing than other aspects of VA administrative data. Findings from prior studies indicate that those with missing data were most often white and had low healthcare utilization and disease burden. For these demographic variables, a process was used in which VA data for previous years was looked back at and data in subsequent years was looked forward at to minimize missing data.

Health Status Variables

Health status variables in the analyses predicting HEDIS HRME exposure in FY06 included several measures of disease burden. Because prior studies indicated that individuals with greater disease burden as defined by more medications, more physical comorbidities and psychiatric conditions are at greater risk for potentially inappropriate prescribing,[7,16–18] these variables were controlled for. The number of unique medication classes each individual received during FY05 was first counted. *International Classification of Diseases, Ninth Revision, Clinical Modification*, codes found in VA inpatient and outpatient data (diagnoses in two outpatients or one inpatient) (FY03–05) were also used to identify individuals with physical and psychiatric conditions using the Selim comorbidity indices that were developed to control for disease burden in research studies involving veterans.[19,20] For physical conditions, the number of chronic disease states from 30 possible conditions included in the Selim Physical Comorbidity index was counted. The following psychiatric conditions included in the Selim Psychiatric Comorbidity Index were also identified: schizophrenia, bipolar disorder, depressive disorder, posttraumatic stress disorder, substance use disorder, and anxiety disorders. Because of the highly skewed distribution, individuals with zero, one, or two or more psychiatric conditions were identified. These measures of comorbidity have been previously associated with mortality, measures of health status, and PIPE.[14,21,22]

Access-to-Care Variables

Access-to-care variables in the analyses included a measure of copayment exemption and measures of geriatric care and primary care utilization in FY05. Copayment exemption was measured using the VA priority group. VA priority groups are associated with physical or mental health status and illness severity and socioeconomic status. Veterans with a service-connected disability of 50% or greater, or individuals who were catastrophically disabled, had very

Table 1. High-Risk Medications in Older Adults

Drug Group	Drugs Included	Concerns Regarding Use
Amphetamines	Amphetamine/dextroamphetamine, benzphetamine, dexmethylphenidate, pemoline, dextroamphetamine diethylpropion, methamphetamine, methylphenidate, phendimetrazine, phentermine	Dependence, hypertension, angina pectoris, and myocardial infarction
Antibiotics	Nitrofurantoin	Potential for renal impairment; safer alternatives available
Antihistamines	Diphenhydramine	Confusion and sedation
	hydroxyzine, promethazine, cyproheptadine, dexchlorpheniramine, tripelennamine	Potent anticholinergic properties
Cardiac medications and vasodilators	Dipyridamole (short acting)	Risk of orthostatic hypotension
	Nifedipine (short acting)	Potential for hypotension and constipation
	Isoxsuprine, ergot mesyloids	Lack of efficacy
Endocrine drugs	Estrogens	Carcinogenic potential
	Chlorpropamide	Syndrome of inappropriate antidiuretic hormone secretion; hypoglycemia
	Desiccated thyroid	Concerns about cardiac effects; safer alternatives available
Gastrointestinal antispasmodics	Dicyclomine, hyoscyamine, propantheline, atropine belladonna, scopolamine	Highly anticholinergic with uncertain effectiveness
	trimethobenzamide	Poor efficacy; can cause extrapyramidal effects
Nonsteroidal anti-inflammatory drugs	Ketorolac	Commonly causes asymptomatic gastrointestinal pathological conditions
Opioid pain relievers	Propoxyphene	Lack of efficacy; more adverse effects
	Meperidine	Lack of efficacy; confusion, falls, fractures, dependency
	Pentazocine	Falls, fractures, confusion, dependency, withdrawal
Psychotropic drugs	Diazepam, chlordiazepoxide, flurazepam	Prolonged sedation and increase the risk of falls
	Thioridazine, meprobamate	More central nervous system and extrapyramidal adverse effects than others
	Barbiturates	Highly addictive; more adverse effects than others
Skeletal muscle relaxants	Methocarbamol, cyclobenzaprine, carisoprodol, chlorzoxazone, metaxalone, orphenadrine	Anticholinergic adverse effects, sedation, and weakness; effectiveness at doses tolerated by older adults questionable

low income, or had specific war-related experiences generally received a waiver for copayments associated with VA care.[21,23]

Because prior work found that geriatric care at some point the year before assessment was associated with lower risk of exposure to drugs included in the Beers criteria, individuals who received care in geriatric outpatient clinics or inpatient geriatric evaluation and management in FY05 were identified as having prior geriatric care.[24] Finally, because prior literature found that patients with many primary care visits the previous year were more likely to have an exposure to potentially inappropriate medications as measured by the Beers criteria, those with more-frequent visits to primary care ($\geq$5 in a year) may be sicker and thus at higher risk of HEDIS HRME exposure.[16,17] Based on prior studies and the empirical distribution, patients were classified as having 0 to 1, 2 to 4, or 5 or more primary care visits.

Analysis

The demographic and health status characteristics of the cohort are first described, and then changes in exposure to HEDIS HRME overall and individual medication classes of older VA outpatients over the 3-year study period (FY04-06) were identified. Generalized estimating equation (GEE) analyses with a logit link (exchangeable working correlation) were used and applied to five unique samples of 100,000 randomly selected individual to determine whether changes over time in HEDIS HRME overall and drug classes were statistically significant. The averaged parameter estimates (change over time only) obtained from the five random samples were used, and standard errors were approximated for the entire population using the pooled standard errors.[25] Statistical significance of these pooled estimates was examined using the *Z*-score threshold: $\pm$ 2.33 ($-$ 2.33 and 2.33 are 0.005 and 0.995 quantiles of the standard normal distribution). Logistic regression analyses were then used to identify demographic, health status, and access-to-care factors associated with risk of HEDIS HRME exposure overall in 2006 and for the four most commonly used HEDIS HRME drug classes. All statistical analyses were performed using SAS software (version 9, SAS Institute, Inc., Cary, NC).

RESULTS

Of the 1,933,291 individuals who met age criteria in FY04, 1,567,467 received care between FY03 and FY06 and were included in this study. The mean age of individuals in this cohort in 2004 was 74.4 $\pm$ 5.8. Similar to other studies of older veterans, this cohort was primarily male (1,539,324, 98.2%) and white (1,060,366, 67.7%). Table 2 provides additional descriptive statistics for this cohort.

Longitudinal Change in HEDIS HRME Exposure

Overall, the rates of HEDIS HRME exposure decreased over the study period (FY04, 205,179, 13.1%; FY05, 200,326, 12.8%; FY06, 193,456, 12.3%). This represents an absolute difference of 0.8% and a relative difference of 6.1% between FY04 and FY06. Figure 1 shows changes in HEDIS HRME exposure according to drug category over the study period. Most categories experienced small reduc-

Table 2. Characteristics of Longitudinal Cohort

Characteristic	Overall (N = 1,567,467)	No HEDIS HRME Exposure (n = 1,374,016)	HEDIS HRME Exposure (n = 193,451)
Age, mean ± SD	74.4 (5.8)	76.5 (5.8)	75.9 (5.8)
Sex, n (%)			
Male	1,539,324 (98.2)	1,352,233 (98.4)	187,091 (96.71)
Female	28,143 (1.8)	21,783 (1.6)	6,360 (3.3)
Race, n (%)			
White	1,060,366 (67.7)	919,575 (66.9)	140,791 (72.8)
African American	103,818 (6.6)	87,261 (6.4)	16,557 (8.6)
Hispanic	52,925 (3.4)	43,913 (3.2)	9,012 (4.7)
Other	19,719 (1.3)	16,988 (1.2)	2,731 (1.4)
Missing	330,639 (21.09)	306,279 (22.3)	24,360 (12.6)
Unique medications, mean (± SD)	6.4 (4.4)	6.1 (4.1)	9.2 (5.1)
Selim Physical Comorbidity Index, mean (± SD)	2.5 (1.7)	2.4 (1.6)	3.1 (1.9)
Selim Psychiatric Comorbidity Index, n (%)			
0	1,362,490 (86.9)	1,212,466 (88.2)	150,024(77.5)
1	159,426 (10.2)	127,866 (9.3)	31,560 (16.3)
≥2	45,551 (2.9)	33,684 (2.5)	11,867 (6.1)
Copayment status, n (%)*			
Exempt	953,467(60.8)	805,186 (58.6)	148,281 (76.7)
Not exempt	613,810 (39.2)	568,662 (41.4)	45,148 (23.3)
Geriatric care in 2003, n (%)			
Yes	33,046 (2.11)	29,020 (2.1)	4,026 (2.1)
No	1,534,421 (97.9)	1,344,996 (97.9)	189,425 (97.9)
Number of primary care visits, n (%)			
<2	382,741 (24.4)	352,852 (25.7)	29,889 (15.5)
2–4	892,662 (57.0)	787,111 (57.3)	105,551 (54.7)
≥5	292,064 (18.6)	234,053 (17.0)	58,011 (30.0)

* 190 individuals had missing data for copayment status.
HEDIS HRME = high-risk medications for the elderly defined according to the Healthcare Effectiveness Data and Information Set quality measure; SD = standard deviation.

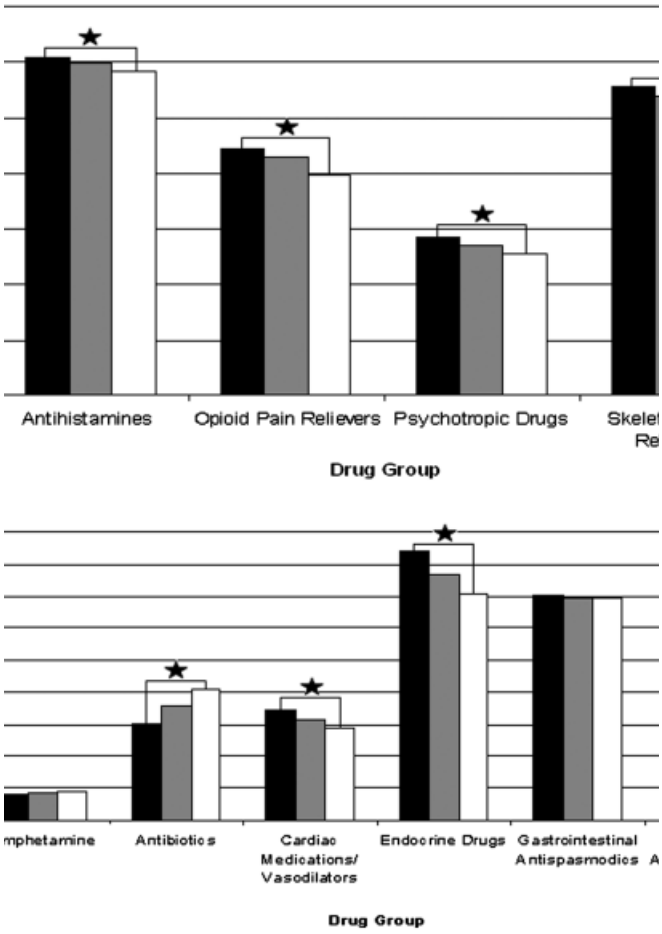


Figure 1. Exposure to high-risk medications in the elderly (HRME) according to drug group.
Trends in use of HRME: 2004 to 2006.
Significant at $P < .01$.
Changes in the number of older veterans exposed to specific high-risk medication classes between FY04 and FY06. Different scales are used for the most commonly used high-risk medication classes and those less commonly used to visualize the change in use over time: [■] FY04; [▩] FY05; [□] FY06

tions in use between FY04 and FY06, although there were relative increases of 36.5%, 10.3%, and 8.0% for antibiotics, amphetamines, and ketorolac (a nonsteroidal anti-inflammatory drug), respectively. Table 3 shows results of GEE analyses examining the statistical significance of HEDIS HRME over time in the five random samples of the population. Overall, exposure to HEDIS HRME was significantly lower in FY06 than FY04 (estimate −0.07, standard error 0.004, *Z*-score −15.59). There was no significant change in exposure for gastrointestinal antispasmodics, amphetamines, or ketorolac, but there were statistically significant reductions in high-risk opioid pain relievers (primarily propoxyphene), skeletal muscle relaxants, psychotropic drugs, endocrine drugs, and cardiac and vasodilator medications. There was also a statistically significant increase in exposure to nitrofurantoin.

Table 4 shows results of logistic regression models predicting HEDIS HRME exposure in FY06. Similar to a previous report using data from FY00, older individuals, African Americans (vs whites), those required to make copayments, and those with previous geriatric care were significantly less likely to have HEDIS HRME exposure. Women; Hispanics; and those with higher numbers of medications the prior year, psychiatric comorbidity, or higher primary care utilization were significantly more likely to have HEDIS HRME exposure.

Examination of the four types of the most commonly used HEDIS HRME suggested some consistency across medication types but also some variation with regard to predictors of HEDIS HRME exposure (Table 4). The effect of age, sex, number of medications, copayment requirement, and prior geriatric care were consistent across all drug groups. African Americans and Hispanics were less likely than whites to have opioid HEDIS HRME exposure and more likely to have antihistamine and skeletal muscle

Table 3. Results of Pooled General Estimating Equations Assessing Change in Proportion of Exposure to Healthcare Effectiveness Data and Information Set (HEDIS) High-Risk Medications in the Elderly (HRME) According to Drug Category FY06 (vs FY04)

Drug Group	Estimate	Standard Error	*Z* Score
Any HEDIS HRME	−0.07	0.004	−15.59*
Amphetamine	0.10	0.045	2.09
Antibiotic	0.30	0.026	11.55*
Antihistamine	−0.04	0.008	−4.40*
Cardiac medication or vasodilator	−0.16	0.018	−8.84*
Endocrine drug	−0.18	0.013	−14.05*
Gastrointestinal antispasmodic	−0.026	0.016	−1.60
Nonsteroidal anti-inflammatory drug	0.03	0.064	0.49
Opioid pain reliever	−0.13	0.009	−13.92*
Psychotropic drug	−0.11	0.009	−12.47*
Skeletal muscle relaxant	−0.08	0.009	−9.19*

* $P < .01$.

relaxant HEDIS HRME exposure. African Americans were also significantly less likely than whites to have psychotropic HEDIS HRME exposure, whereas Hispanics were significantly more likely to have such exposure. Individuals who were not married were more likely to have antihistamine and opioid HEDIS HRME exposure and less likely to have psychotropic or skeletal muscle relaxant HEDIS HRME exposure. The effect of high primary care utilization was associated with greater likelihood of HEDIS HRME exposure, with the exception of psychotropics, for which the effect was not significant for those with two to four visits and significantly greater for those with five or more visits. Finally, psychiatric comorbidity was not significantly associated with exposure to opioids or musculoskeletal relaxants.

DISCUSSION

Findings from this study suggest that small, but statistically significant reductions in exposure to HRME overall defined using the HEDIS criteria occurred between 2004 and 2006 for older VA patients. Although the reductions in overall exposure were small during this time period (13.1% in FY04 to 12.3% in FY06), exposure was markedly lower than in a previous assessment of 19.6% in FY00 and similar to the rate of 12.9% reported in a previous national sample of retirees who had worked for the same company in 2003 to 2005.[26] The rates of exposure in the current study were considerably lower than those reported by the NCQA for 2006 among Medicare enrollees, with 23.1% of individuals meeting criteria for the denominator in 2006 having one or more HEDIS HRME exposures.[15] The VA's leadership in geriatric care, the active role of pharmacists in the VA, and VA formulary management may have contributed to the reduction. The VA created its One-National Formulary in

Table 4. Logistic Regression Analysis Predicting Exposure to High-Risk Medications for the Elderly (HRME)

	Odds Ratio (99% Confidence Interval)				
Variable	Overall HEDIS HRME	Antihistamines	Opioids	Psychotropics	Skeletal Muscle Relaxants
Demographic					
Age	0.98 (0.98–0.98)	0.99 (0.99–0.99)	1.00 (1.00–1.00)	0.98 (0.98–0.98)	0.94 (0.94–0.94)
Female (vs male)	1.95 (1.87–2.02)	1.16 (1.11–1.20)	1.29 (1.17–1.41)	1.23 (1.10–1.36)	1.44 (1.33–1.56)
Race (vs white)					
African American	0.96 (0.93–0.98)	1.16 (1.11–1.20)	0.73 (0.69–0.77)	0.75 (0.70–0.80)	1.13 (1.09–1.18)
Hispanic	1.08 (1.05–1.12)	1.34 (1.28–1.41)	0.32 (0.28–0.35)	1.10 (1.02–1.19)	1.41 (1.34–1.48)
Other	0.97 (0.92–1.03)	1.18 (1.08–1.28)	0.69 (0.60–0.78)	0.97 (0.85–1.12)	0.98 (0.89–1.08)
Missing	0.77 (0.75–0.78)	0.67 (0.65–0.70)	0.66 (0.64–0.69)	0.88 (0.84–0.93)	0.78 (0.75–0.81)
Not married (vs married)	1.00 (0.99–1.02)	1.14 (1.12–1.17)	0.92 (0.90–0.95)	0.88 (0.85–0.91)	0.97 (0.95–1.00)
Health status factors					
Unique medications	1.13 (1.12–1.13)	1.12 (1.11–1.12)	1.10 (1.09–1.10)	1.12 (1.12–1.13)	1.10 (1.10–1.10)
Selim Physical Comorbidity Index	0.99 (0.98–0.99)	0.96 (0.95–0.97)	1.07 (1.06–1.08)	0.89 (0.88–0.90)	1.01 (1.00–1.02)
Selim Psychiatric Comorbidity Index (vs 0)					
1	1.38 (1.35–1.40)	1.35 (1.31–1.40)	1.03 (0.99–1.08)	2.79 (2.68–2.91)	1.06 (1.03–1.10)
2 or more	1.51 (1.47–1.56)	1.53 (1.46–1.60)	0.92 (0.86–0.99)	3.75 (3.54–3.96)	1.06 (1.00–1.12)
Access-to-care factors					
Copayment status not exempt	0.61 (0.60–0.62)	0.49 (0.47–0.50)	0.61 (0.59–0.63)	0.56 (0.54–0.59)	0.55 (0.53–0.56)
Received geriatric care	0.73 (0.70–0.77)	0.71 (0.66–0.77)	0.69 (0.63–0.76)	0.65 (0.57–0.73)	0.65 (0.59–0.71)
Number of primary care visits (vs <2)					
2–4	1.10 (1.08–1.12)	1.10 (1.07–1.14)	1.20 (1.16–1.25)	1.00 (0.96–1.05)	1.24 (1.20–1.29)
≥5	1.12 (1.09–1.15)	1.20 (1.15–1.25)	0.99 (0.95–1.04)	0.87 (0.82–0.92)	1.35 (1.30–1.41)

HEDIS = Healthcare Effectiveness Data and Information Set.

2002 by freezing formularies at the facility and regional level and then creating a uniform formulary, which resulted in excluding drugs that had previously been on a number of regional formularies, such as propoxyphene.

Examination of change over time according to individual drugs and drug classes revealed some reduction in most drug classes, stability in others, and increases in nitrofurantoin. Reductions in use of skeletal muscle relaxants, psychotropic drugs, and opioid pain relievers is notable because recent studies have demonstrated that the use of these medications increases the risk of falls and fractures in older adults.[27–29]

One possible explanation for the increase in nitrofurantoin use is related to increasing resistance to common urinary tract pathogens such as *Escherichia coli* with common antibiotics (e.g., ciprofloxacin, trimethoprim/sulfamethoxazole). One in vitro study found that nitrofurantoin was effective in killing *E. coli* isolates in 98.1% of those with trimethoprim/sulfamethoxazole resistance and 89.6% of those with ciprofloxacin resistance.[30] Unfortunately, in vitro testing does not translate to nitrofurantoin being effective in older adults. The use of nitrofurantoin, a primarily renally cleared medication, should be avoided in individuals with estimated creatinine clearances less than 60 mL/min because insufficient concentrations reach the bladder to be capable of killing bacteria such as *E. coli*.[31,32] A recent study of veterans residing in a VA community living center found that this agent was in the top four suboptimally prescribed medications.[33] Of concern is a greater risk for serious adverse drug events with nitrofurantoin that include chronic, subacute, or acute pulmonary hypersensitivity reactions and peripheral neuropathy.

This study also adds to the understanding of risk factors for potentially inappropriate prescribing. Prior studies have examined a single drug such as propoxyphene or potentially inappropriate drugs in the Beers or HEDIS criteria as a single entity.[7,14–18,34,35] Neither of the aforementioned studies examined risk factors for the use of high-risk medications.[15,26] Consistent with a prior report and other studies examining exposure to potentially inappropriate medications, it was found that whites, women, and those with more medications were more likely to be exposed.[7,18] Examination of the four most commonly prescribed HEDIS HRME groups suggests that findings from studies of HEDIS HRME as a whole provide insufficient insight into this problem. In particular, the effects of race and psychiatric comorbidity and primary care utilization depend upon the type of potentially inappropriate medication.

With regard to race, African Americans were less likely to have exposure to suboptimal opioid and psychotropic medications than were whites, and Hispanics were also significantly less likely than whites to have exposure to opioid medications. The finding for African Americans is consistent with literature finding lower use of psychotropic medications and analgesics in blacks than in whites.[36,37] The finding for Hispanics is less clear, in part because many previous studies have not distinguished between blacks and Hispanics but rather evaluated them as "nonwhites."

Individuals with multiple psychiatric comorbidities had a lower risk of being prescribed high-risk opioids (e.g., propoxyphene). This finding may result from clinicians being less likely to prescribe opioids for pain in individuals receiving psychotropic medications for psychiatric comorbidities because of concern of that greater total central nervous system medication burden (e.g., opioids, benzodiazepines, antidepressants, antipsychotics) increase the risk of falls in older adults.[38]

A number of potential limitations should be noted. First, restriction of the assessment to individuals who received VA care between FY03 and FY06 may lead to selection bias. Although this was necessary to understand change in a consistent population, this may bias the results because individuals who were sicker and died during this 3-year period are not represented in the findings, but examination of the entire population revealed similar rates of exposure and trends overall and for HEDIS HRME drug groups (range 14.1% in FY04 to 12.6% in FY06) and predictors of exposure. Second, the assessment was restricted to medications received within the VA. It is possible that some HEDIS HRME were purchased outside the VA, and thus the assessment may be conservative. One potential problem is with medications that can be purchased over the counter (OTC), such as diphenhydramine. Substantial variation in the relationship between copayment status and HEDIS HRME exposure for antihistamines would support the idea that OTC medications affected lower risk for exposure differently in those with required copayments. Although small variations existed, the direction and magnitude of the copayment variable was similar across drug classes. A second concern regarding use of medications received from the VA is that implementation of Medicare Part D in January 2006 may have affected the findings.[39] Because a sudden, marked decrease in the average number of prescriptions per patient or marked decreases in exposure between 2005 and 2006 were not seen (Table 2), it is unlikely that the findings for FY06 were affected substantially. Moreover, the assessment occurred using data from a timeframe before the implementation of the HEDIS HRME measure. Because the Beers criteria from which the HEDIS HRME measure was derived have been in existence in various forms for nearly a decade, the time period examined for this assessment is not unreasonable. Moreover, this study provides a foundation for subsequent study of change in exposure by chronicling the years up to and including the first year of HEDIS HRME implementation. Residual confounding due to potential important factors for which information was not available (e.g., smoking) and HEDIS HRME cannot be excluded. Because this study provides information on change in prescribing on a national sample of older primarily male VA patients, it does not reflect prescribing in non-VA settings. Finally, because propoxyphene was removed from the market, it is expected that rates of HEDIS HRME exposure will fall significantly nationally. This would follow a consistent trend in the VA, where rates fell considerably due to formulary restrictions, but because opioids rank third among the most common high-risk drug classes, high-risk medication exposure continues to be of concern.

CONCLUSION

This study found a small decrease in HEDIS HRME exposure between FY04 and FY06. These rates of exposure were lower in the VA than in 2006 Medicare data. Comparison

of the findings with those of an earlier study[14] suggests that exposure in the VA has fallen substantially between 2000 and 2006 but that only small changes occurred between 2004 and 2006. Moreover, variation in risk factors was found for different groups of HEDIS HRME drugs. Future studies should examine the effect of these reductions on overall health outcomes and measure the effects according to drug class because variation in outcomes may also be evident, depending on the medications and the conditions they treat.

ACKNOWLEDGMENTS

The authors acknowledge and appreciate support from the South Texas Veterans Healthcare System/Audie L. Murphy Division and the Veterans Evidence-based Research and Implementation Center. The views expressed in this article are those of the authors and do not necessarily represent the views of the VA. We acknowledge the contributions of Jeffrey Tabares and Ranjani Davalath in assisting with manuscript preparation.

Conflict of Interest: The editor in chief has reviewed the conflict of interest checklist provided by the authors and has determined that the authors have no financial or any other kind of personal conflicts with this paper.

Dr. Pugh has received research funding from VA Health Services Research and Development Service (HSRD) DHI-09-237 (PI); VA HSRD IIR-06-062 PI, Epilepsy Foundation PI, VA HSRD PPO 09-295 PI, VA HSRD IIR 02-274 PI. Pugh as co-I: VA HSRD IIR 08-274, VA HSRD SDR-07-042, IIR-05-121, IAF-06-080, IIR-09-335, SHP 08-140, TRX 01-091 Department of Defense CDMRP 09090014, NIH R01-NR010828,

This study was supported by a VA Health Services Research grant (IIR-06–062, Dr. Pugh PI). Dr Hanlon is supported by National Institute of Aging grants (R01AG027017, P30AG024827, T32 AG021885, K07AG033174, R01AG034056), a National Institute of Mental Health grant (R34 MH082682), a National Institute of Nursing Research grant (R01NR010135), and an Agency for Healthcare Research and Quality grant (R01HS017695).

Author Contributions: Mary Jo V. Pugh: concept and design, acquisition of data, analysis and interpretation of data and preparation of the manuscript. Dan R. Berlowitz, Todd Semla, Muriel Burk, and John R. Downs: interpretation of data and preparation of the manuscript. Megan E. Amuan: acquisition of data, analysis of data and preparation of the manuscript. Chen-Pin Wang: analytical design, analysis and interpretation of data and preparation of the manuscript. Chester B. Good: interpretation of data and preparation of the manuscript. Joseph T. Hanlon: study design, interpretation of data and preparation of the manuscript.

Sponsor's Role: VA HSRD played no role in the design, methods, subject recruitment, data collections, analysis, or preparation of paper.

REFERENCES

1. Fick DM, Waller JL, Maclean JR et al. Potentially inappropriate medication use in a Medicare managed care population with higher costs and utilization. J Manage Care Pharm 2001;7:407–413.
2. Zhan C, Correa-de-Araujo R, Bierman AS et al. Suboptimal prescribing in elderly outpatients: Potentially harmful drug-drug and drug-disease combinations. J Am Geriatr Soc 2005;53:262–267.
3. Blalock SJ, Byrd JE, Hansen RA et al. Factors associated with potentially inappropriate drug utilization in a sample of rural community-dwelling older adults. Am J Geriatr Pharmacother 2005;3:168–179.
4. Cannon KT, Choi MM, Zuniga MA. Potentially inappropriate medication use in elderly patients receiving home health care: A retrospective data analysis. Am J Geriatr Pharmacother 2006;4:134–143.
5. Barnett MJ, Perry PJ, Langstaff JD et al. Comparison of rates of potentially inappropriate medication use according to the Zhan criteria for VA versus private sector medicare HMOs. J Manage Care Pharm 2006;12:362–370.
6. Roughead EE, Anderson B, Gilbert AL. Potentially inappropriate prescribing among Australian veterans and war widows/widowers. Intern Med J 2007;37:402–405.
7. Pugh MJ, Fincke BG, Bierman A et al. Potentially inappropriate prescribing in elderly veterans: Are we using the wrong drug, wrong dose, or wrong duration? J Am Geriatr Soc 2005;53:1282–1289.
8. Shrout PE, Yager TJ. Reliability and validity of screening scales: Effect of reducing scale length. J Clin Epidemiol 1989;42:69–78.
9. Stuart BC, Kamal-Bahl SJ, Briesacher BA et al. Trends in prescribing inappropriate drugs for the elderly between 1995 and 1999. Am J Geriatr Pharmacother 2003;1:61–74.
10. Bongue B, Naudin F, Laroche ML et al. Trends of the potentially inappropriate medication consumption over 10 years in older adults in the east of France. Pharmacoepidemiol Drug Saf 2009;18:1125–1133.
11. Carey IM, De Wilde S, Harris T et al. What factors predict potentially inappropriate primary care prescribing in older people? Analysis of UK Primary Care Patient Record database. Drugs Aging 2008;25:693–706.
12. Use of High Risk Medications in the Elderly. National Committee on Quality Assurance [on-line]. Available at http://www.ncqa.org/portals/0/hedisqm/HEDIS2008/Vol2/NDC/Table%20DAE-A.doc Accessed October 24, 2005.
13. Fick DM, Cooper JW, Wade WE et al. Updating the Beers criteria for potentially inappropriate medication use in older adults: Results of a US consensus panel of experts. Arch Intern Med 2003;163:2716–2724.
14. Pugh MJ, Hanlon JT, Zeber JE et al. Assessing potentially inappropriate prescribing in the elderly Veterans Affairs population using the HEDIS 2006 quality measure. J Manage Care Pharm 2006;12:537–545.
15. The state of health care quality: Value, variation and vulnerable populations 2009. National Committee for Quality Assurance [on-line]. Available at http://www.ncqa.org/Portals/0/Newsroom/SOHC/SOHC_2009.pdf Accessed August 6, 2010.
16. Zhan C, Sangl J, Bierman AS et al. Potentially inappropriate medication use in the community-dwelling elderly: Findings from the 1996 Medical Expenditure Panel Survey. JAMA 2001;286:2823–2829.
17. Hanlon JT, Fillenbaum GG, Schmader KE et al. Inappropriate drug use among community dwelling elderly. Pharmacotherapy 2000;20:575–582.
18. Goulding MR. Inappropriate medication prescribing for elderly ambulatory care patients. Arch Intern Med 2004;164:305–312.
19. Selim AJ, Berlowitz DR, Ren XS et al. The comorbidity index. In: Davies M, editor. Measuring and Managing Health Care Quality, Vol. 4. New York: Aspen Publishers, 2002, pp 91–94.
20. Selim AJ, Fincke G, Ren XS et al. Comorbidity assessments based on patient report: Results from the Veterans Health Study. J Ambulatory Care Manage 2004;27:281–295.
21. Pugh MJ, Copeland LA, Zeber JE et al. The impact of epilepsy on health status among younger and older adults. Epilepsia 2005;46:1820–1827.
22. Selim AJ, Kazis LE, Rogers W et al. Risk-adjusted mortality as an indicator of outcomes: Comparison of the medicare advantage program with the Veterans' Health Administration. Med Care 2006;44:359–365.
23. Kazis LE, Miller DR, Clark J et al. Health-related quality of life in patients served by the Department of Veterans Affairs. Arch Intern Med 1998;158:626–632.
24. Pugh MJ, Rosen AK, Montez-Rath M et al. Potentially inappropriate prescribing for the elderly: Effects of geriatric care at the patient and health care system level. Med Care 2008;46:167–173.
25. Hanley JA, Negassa A, Edwardes MD et al. Statistical analysis of correlated data using generalized estimating equations: An orientation. Am J Epidemiol 2003;157:364–375.
26. Albert SM, Colombi A, Hanlon J. Potentially inappropriate medications and risk of hospitalization in retirees: Analysis of a US retiree health claims database. Drugs Aging 2010;27:407–415.
27. Woolcott JC, Richardson KJ, Wiens MO et al. Meta-analysis of the impact of 9 medication classes on falls in elderly persons. Arch Intern Med 2009;169:1952–1960.
28. Hartikainen S, Lonnroos E, Louhivuori K. Medication as a risk factor for falls: Critical systematic review. J Gerontol A Biol Sci Med Sci 2007;62A:1172–1181.

29. Golden AG, Ma Q, Nair V et al. Risk for fractures with centrally acting muscle relaxants: An analysis of a National Medicare Advantage claims database. Ann Pharmacother 2010;44:1369–1375.
30. Karlowsky JA, Thornsberry C, Jones ME et al. Susceptibility of antimicrobial-resistant urinary *Escherichia coli* isolates to fluoroquinolones and nitrofurantoin. Clin Infect Dis 2003;36:183–187.
31. Sachs J, Geer T, Noell P et al. Effect of renal function on urinary recovery of orally administered nitrofurantoin. N Engl J Med 1968;278:1032–1035.
32. Covvey JR, Lewis DA. Glimepiride-induced hypoglycemia with ciprofloxacin, metronidazole, and acute kidney injury. Hosp Pharm 2010;45:934–938.
33. Hanlon JT, Wang X, Handler SM et al. Potentially inappropriate prescribing of primarily renally cleared medications for older Veterans Affairs nursing home patients. J Am Med Dir Assoc 2011;12:377–383.
34. Kamal-Bahl S, Stuart BC, Beers MH. National trends in and predictors of propoxyphene use in community-dwelling older adults. Am J Geriatr Pharmacother 2005;3:186–195.
35. Simon SR, Chan KA, Soumerai SB et al. Potentially inappropriate medication use by elderly persons in U.S. Health Maintenance Organizations, 2000–2001. J Am Geriatr Soc 2005;53:227–232.
36. Hanlon JT, Fillenbaum GG, Studenski SA et al. Factors associated with suboptimal analgesic use in community-dwelling elderly. Ann Pharmacother 1996;30:739–744.
37. Blazer D, Hybels C, Simonsick E et al. Sedative, hypnotic, and antianxiety medication use in an aging cohort over ten years: A racial comparison. J Am Geriatr Soc 2000;48:1073–1079.
38. Hanlon JT, Boudreau RM, Roumani YF et al. Number and dosage of central nervous system medications on recurrent falls in community elders: The Health, Aging and Body Composition Study. J Gerontol A Biol Sci Med Sci 2009;64A: 492–498.
39. Resources on the Medicare Prescription Drug Benefit. Kaiser Family Foundation [on-line]. Available at http://www.kff.org/medicare/rxdrugbenefit.cfm Accessed October 16, 1999.

Exposure to Potentially Harmful Drug–Disease Interactions in Older Community-Dwelling Veterans Based on the Healthcare Effectiveness Data and Information Set Quality Measure: Who Is at Risk?

Mary Jo V. Pugh, PhD,[abc] Catherine I. Starner, PharmD,[d] Megan E. Amuan, MPH,[e] Dan R. Berlowitz, MD, MPH,[ef] Monica Horton, MD,[cg] Zachary A. Marcum, PharmD,[h] and Joseph T. Hanlon, PharmD, MS[hijkl]

OBJECTIVES: To identify prevalence and risk factors for exposure to drug–disease interactions included in the Healthcare Effectiveness Data and Information Set (HEDIS) Drug–Disease Interaction (Rx-DIS) measure.

DESIGN: Cross-sectional retrospective database analysis.

SETTING: Outpatient clinics within the Department of Veterans Affairs (VA).

PARTICIPANTS: Individuals aged 65 and older who received VA outpatient care between October 1, 2003, and September 30, 2006.

MEASUREMENTS: Rx-DIS exposure based on the HEDIS measure was identified in VA patients with dementia, falls, and chronic renal failure using VA pharmacy and administrative databases. Factors associated with Rx-DIS exposure were examined, including demographic, health status, and access-to-care factors, including VA outpatient health services use and copayment status.

RESULTS: Of the 305,041 older veterans who met criteria for inclusion, the 1-year prevalence of Rx-DIS exposure was 15.2%; prevalence was 20.2% for dementia, 16.2% for falls, and 8.5% for chronic renal failure. Patients with high disease burden (physical, psychiatric, number of medications) were significantly more likely to have Rx-DIS exposure, regardless of condition. Hispanics and individuals with no copayments were more likely to have Rx-DIS exposure than whites or those with required copayments. There was variation in other predictors based on the type of Rx-DIS.

CONCLUSION: The prevalence of Rx-DIS was common in older VA outpatients. Future studies should examine the risk of Rx-DIS exposure on health outcomes using separate analyses for each type of Rx-DIS separately before combining all Rx-DIS into a single measure of exposure. Studies that examine the effectiveness of interventions to reduce Rx-DIS exposure will also be helpful in improving the quality of care for older adults. **J Am Geriatr Soc 59:1673–1678, 2011.**

Key words: drug disease interaction; HEDIS measures; potentially inappropriate prescribing; aged; pharmacoepidemiology

From the [a]Veterans Evidence-based Research and Implementation Center and [g]Geriatrics and Extended Care, South Texas Veterans Health Care System, San Antonio, Texas; [b]Department of Epidemiology and Biostatistics and, [c]Division of Geriatrics, Gerontology, and Palliative Medicine Health Science Center, University of Texas Health Science Center at San Antonio, San Antonio, Texas; [d]Prime Therapeutics, Minneapolis, Minnesota; [e]Edith Nourse Rogers Memorial Veterans Affairs Hospital, Center for Health Quality Outcomes and Economic Research, Bedford, Massachusetts; [f]School of Public Health, Boston University, Boston, Massachusetts; [h]Division of Geriatrics, Department of Medicine, School of Medicine, [i]Department of Pharmacy and Therapeutics, School of Pharmacy, [j]Department of Epidemiology, Graduate School of Public Health, and [k]Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, Pennsylvania; and [l]Center for Health Equity Research and Geriatric Research, Education and Clinical Center, Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, Pennsylvania.

An abstract presenting a portion of these data was presented at the Annual Meeting of the American Geriatrics Society, Orlando, Florida, May 2010.

Address correspondence to Mary Jo Pugh, South Texas Veterans Health Care System, VERDICT REAP (11C6), 7400 Merton Minter, San Antonio, TX 78229. E-mail: pughm@uthscsa.edu or maryjo.pugh2@va.gov

DOI: 10.1111/j.1532-5415.2011.03524.x

Potentially inappropriate prescribing in the elderly (PIPE) has been a growing concern over the past decade. Although most studies have examined use of high-risk drugs for older adults (e.g., Beers criteria), concern has begun to expand to other realms of PIPE such as drug–disease interactions.[1–5] Studies have examined exposure to drug–disease interactions defined by Beers in a variety of settings.[1,2,6–13] Previous research has shown that drug–disease interactions (medication(s) exacerbating preexisting conditions) are common and are associated with adverse drug reactions in older adults, thus representing an important area of inquiry.[14,15]

0002-8614/11/$15.00

The National Committee on Quality Assurance (NCQA) developed a drug–disease interaction measure as part of the 2007 Healthcare Effectiveness Data and Information Set (HEDIS) Drug–Disease Interaction (Rx-DIS) measure based on an earlier measure of 28 drug–disease interactions involving 14 diseases or conditions developed by Lindblad and colleagues (e.g., peptic ulcer disease and aspirin and nonaspirin nonsteroidal anti-inflammatory drugs (NSAIDs), syncope and alpha blockers, and systolic heart failure and first-generation calcium channel blockers).[7] From the Lindblad measure, the NCQA expert panel reached consensus on a subset of drug–disease interactions that could be readily measured using administrative data and that were potentially associated with adverse outcomes. The three conditions and medication groups considered inappropriate for individuals with those conditions are included in the HEDIS Rx-DIS measure that is now used to monitor quality of prescribing in older patients diagnosed with dementia, falls, and chronic renal failure. Although NCQA has published rates of HEDIS Rx-DIS in their report on the state of healthcare quality in 2009,[16] other studies examining Rx-DIS have used broader measures. Because the HEDIS Rx-DIS measure is a nationally accepted quality measure, the focus of the current study was assessment of the three conditions included in that measure.

The purpose of this article is to examine the extent to which HEDIS Rx-DIS exposure occurs in older community-dwelling VA patients and factors associated with that exposure. Mirroring the HEDIS Rx-DIS quality measurement, the prevalence of HEDIS Rx-DIS exposure was examined overall and then according to disease or condition. To determine whether risk factors were consistent across conditions, risk factors for HEDIS Rx-DIS exposure were also identified overall and according to disease or condition.

METHODS

Data and Study Population

After institutional review board approval national Veterans Affairs (VA) inpatient, outpatient and pharmacy data were obtained from fiscal year 2004 (FY04; October 1, 2003–September 30, 2004) through FY06 (October 1, 2005–September 30, 2006) for individuals aged 65 and older at the beginning of FY05. Pharmacy and diagnostic datasets were merged using the encrypted identifier included in each data set. To ensure that there were adequate data to identify comorbid conditions and prior medication use, individuals who received care regularly in the VA healthcare system (having at least one outpatient or inpatient visit each year) were selected from that population. Individuals who resided in VA community living centers (previously called nursing home care units; based on VA extended care file data) for all of FY2006 and those who died before 2006 were not included. Individuals who were admitted to a community living center during 2006 or who died after receiving care in 2006 were included. Analyses were further restricted to individuals with International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code–based diagnoses or medications indicative of falls, dementia, or chronic renal failure as outlined by the NCQA, which required a single diagnosis.

Measures

Drug–Disease Interaction

Individuals with dementia, falls, and chronic renal failure in VA inpatient and outpatient databases (FY04–06) were first identified using ICD-9-CM codes (and medications for dementia) identified by the NCQA.[17] For those with a diagnosis indicative of dementia, falls, or chronic renal failure or dementia medications (donepezil, galantamine, rivastigmine, tacrine, and memantine), use of potentially harmful medications in FY06 (Table 1) was identified using the VA product variable in the VA Pharmacy Benefits Management database.

Individuals diagnosed with dementia, falls, or chronic renal failure and who later had an order for and were dispensed medications considered relatively or potentially contraindicated in FY06 were classified as having Rx-DIS exposure for that condition. The overall prevalence of any Rx-DIS in individuals who met criteria for dementia, falls, or chronic renal failure and the prevalence for individuals within each condition of interest were then calculated.

Patient Demographic Characteristics

Patient demographic characteristics included age, sex, race, and marital status. Demographic characteristics were identified using VA administrative data. With the exception of race, these demographic characteristics are well documented and complete in the medical record. Because the process of recording race changed in 2002, race data are more likely to be missing than other aspects of VA administrative data. Thus, a process was used in which VA data for previous and subsequent years was looked at to fill in missing race values. Individuals with missing race data (n = 44,513, 14.6% of the cohort meeting inclusion criteria for quality indicators) were excluded from the analysis

Table 1. Description of Drug–Disease Interactions (Rx-DIS) Identified Using the Healthcare Effectiveness Data and Information Set (HEDIS) Rx-DIS Measure

Disease State (n with Condition or Disease)	Drugs to Avoid	With Condition, n (%)	Exposure in 2006 in National Committee on Quality Assurance Study Using Medicare Data, %[16]
Dementia (131,808)	Anticholinergics,* tricyclic antidepressants	26,640 (20.2)	24.6
Fall or hip fracture (54,393)	Antipsychotics, tricyclic antidepressants, sleep agents	8,806 (16.2)	14.6
Chronic renal failure (154,278)	Nonsteroidal anti-inflammatory drugs	13,165 (8.5)	9.5
≥1 Rx-DIS overall (305,041)		46,481 (15.2)	19.4

* For example, diphenhydramine, dicyclomine, promethazine, cyclobenzaprine, chlorpheniramine, and oxybutynin.

of risk factors for HEDIS Rx-DIS exposure because inclusion of these individuals in a separate missing category complicated interpretation of race findings, and findings were essentially the same in analyses in which they were included.

Health Status Factors

Clinical characteristics included in this study included counts of chronic physical comorbidities, mental health comorbidities, and unique medication classes prescribed the previous year. Preexisting comorbidities were identified between FY04 and FY05 to assure adequate time for accurate assessment.[18] ICD-9-CM codes included in VA inpatient and outpatient databases were used to identify conditions included in the Selim Physical and Mental Comorbidity Indices.[19] The Physical Comorbidity Index includes 30 comorbid conditions (e.g., cerebrovascular disease, diabetes mellitus, hypertension) that are counted to create a physical disease burden score ranging from 0 to 30. The Mental Comorbidity Index includes six comorbid conditions (anxiety, depression, posttraumatic stress disorder, bipolar disorder, alcohol abuse and dependence, and schizophrenia) that are similarly summed to create a score indicating psychiatric disease burden. Previous work has shown that creation of categorical variables assists in interpretation of mental health conditions (0, 1, $\geq$2 conditions). The Physical and Mental Comorbidity Indices have been found to be associated with mortality and suboptimal prescribing in previous research.[5,20] In addition to comorbid conditions, a count of unique medication classes prescribed for each patient in FY05 (the year before Rx-DIS identification) was also used.[21]

Access-to-Care Factors

The first variable measuring access to care was an indicator identifying those for whom pharmacy copayments were required based on VA priority group. VA priority groups are associated with physical and mental health status, illness severity, and socioeconomic status. As defined here, priority group status that warranted a waiver of pharmacy copayments ($8 in 2006) included veterans with a service-connected disability of 50% or greater and individuals who were catastrophically disabled, had very low income, or had specific war-related experiences.[22] The second variable measuring access to care was prior receipt of geriatric care. Individuals who received care in geriatric outpatient clinics or inpatient geriatric evaluation and management in FY05 were identified as having prior geriatric care.[23] The third variable measuring access to care was a count of primary care visits. Because prior research found that patients with many primary care visits the previous year were more likely to have exposure to potentially inappropriate medications as measured according to the Beers criteria, those with more-frequent visits to primary care ($\geq$5 in a year) may be sicker and thus at higher risk of RX-DIS exposure.[3,24] Based on prior studies and the empirical distribution, patients were classified as having 0 or 1, 2 to 4, or 5 or more primary care visits.

Analysis

The prevalence of HEDIS Rx-DIS overall in participants meeting criteria for dementia, falls, and chronic renal failure and the prevalence of HEDIS Rx-DIS exposure according to specific condition is first provided. Then, risk factors for HEDIS Rx-DIS overall and within each condition of interest were identified using logistic regression analysis. Collinearity diagnostic testing was conducted to assure that multicollinearity did not exist between the variables included in the logistic regression models. SAS version 9.1 (SAS Institute, Inc., Cary, NC) was used for all analyses.

RESULTS

Of the 1,780,787 older community-dwelling veterans who received care regularly within the VA each year between FY04 and FY06, 305,041 met criteria for dementia, falls, or chronic renal failure. Table 1 shows the prevalence of drug–disease interactions according to disease state or condition. Based on the HEDIS Rx-DIS criteria, 46,481 (15.2%) of this cohort had one or more Rx-DIS exposures during FY06; 20.2% of the veterans with a history of dementia, 16.2% of those with a history of falls, and 8.5% of those with a history of chronic renal failure had Rx-DIS exposure.

Table 2 shows descriptive statistics for individuals with one or more and no Rx-DIS and for individuals with and without Rx-DIS exposure according to condition. There were statistically significant differences between those with and without Rx-DIS exposure for all bivariate analyses except for marital status in dementia and falls and sex in chronic renal failure. Table 3 shows adjusted odds ratios and 95% confidence intervals from logistic regression models examining the association between Rx-DIS exposure and independent variables overall and for each condition. In the overall analysis, risk factors included age, sex, ethnicity, psychiatric comorbidities, and level of primary care utilization, although there was some variation when examining Rx-DIS exposures within specific disease conditions with regard to race, sex, psychiatric comorbidities, and geriatric and primary care utilization.

DISCUSSION

This study of community-dwelling veterans is the first to report prevalence of exposure to Rx-DIS and potential risk factors for exposure based on the HEDIS measure. The prevalence of Rx-DIS is similar to rates of PIPE as defined according to the Beers criteria and other measures of high-risk medications in older adults (15–40%).[1,8,9,25] Given recent data showing a strong relationship between Rx-DIS exposure and adverse drug reactions, this form of suboptimal prescribing is a serious public health concern.[14,15]

Although other published studies using measures of Rx-DIS exposure that included more conditions and medications had much higher rates of exposure (5.7%,[12,13] 15.3%[26]), in this context, fewer conditions and fewer medications were considered suboptimal. Overall rates of any exposure were lower than population estimates according to the NCQA using the same measure in Medicare data in 2006, but the differences were primarily due to lower Rx-DIS exposure in individuals with dementia and chronic renal failure.[16] Consistent with other studies examining Rx-DIS, the number of unique medications had a strong association with Rx-DIS exposure,[10] although unlike in one study, women and Hispanics (but not men or African Americans) were at greater risk of Rx-DIS overall. These differences in demographic characteristics may be due to the smaller number of

Table 2. Descriptive Statistics: Healthcare Effectiveness Data and Information Set (HEDIS) Drug–Disease Interaction (Rx-DIS) Exposure According to Condition

	Overall		Dementia		Falls		Chronic Renal Failure	
Variable	No (n = 258,560)	Yes (n = 46,481)	No (n = 105,158)	Yes (n = 26,650)	No (n = 45,587)	Yes (n = 8,806)	No (n = 141,113)	Yes (n = 13,165)
Demographic characteristic								
Age, mean ± SD	79.5 ± 5.7	78.5 ± 5.9	79.5 ± 5.8	78.5 ± 5.9	78.7 ± 6.5	77.7 ± 6.4	77.5 ± 6.1	76.3 ± 6.0
Male, n (%)	253,897 (98.2)	42,216 (97.3)	102,832 (97.9)	25,884 (97.1)	43,726 (95.9)	8,371 (95.1)	139,733 (99.0)	13,020 (98.9)
Race, n (%)								
White	176,514 (68.3)	33,017 (71.0)	70,803 (67.4)	18,907 (71.0)	34,574 (75.8)	6,890 (78.3)	94,682 (67.1)	8,802 (66.9)
African American	27,445 (10.6)	5,323 (11.4)	10,399 (9.9)	2,620 (9.8)	4,071 (8.9)	786 (8.9)	17,837 (12.6)	2,176 (16.5)
Hispanic	10,924 (4.2)	2,940 (6.3)	5,432 (5.2)	1,835 (6.9)	2,219 (4.9)	556 (6.3)	5,147 (3.7)	722 (5.5)
Other	3,695 (1.4)	670 (1.4)	1,391 (1.3)	345 (1.3)	697 (1.5)	137 (1.6)	2,122 (1.5)	214 (1.6)
Missing	39,982 (15.5)	4,531 (9.8)	17,054 (16.2)	2,943 (11.0)	4,026 (8.8)	433 (4.9)	21,316 (15.1)	1,251 (9.5)
Married, n (%)	89,680 (34.8)	17.362 (37.5)	69,535 (66.4)	17,427 (65.6)	25,324 (55.7)	4,762 (54.3)	92,087 (65.5)	7,999 (60.9)
Health status factors								
Unique medications, mean ± SD	8.1 ± 4.9	11.6 ± 5.5	8.1 ± 4.9	11.6 ± 5.5	10.3 ± 5.8	13.3 ± 5.9	9.8 ± 5.3	11.6 ± 5.5
Selim Physical, mean ± SD	4.3 ± 2.4	4.9 ± 2.5	3.7 ± 2.3	4.6 ± 2.5	5.2 ± 2.6	5.5 ± 2.7	4.7 ± 2.4	5.2 ± 2.4
Selim Mental, n (%)								
0	192,019 (74.3)	26,367 (56.7)	66,229 (63.0)	13,453 (50.4)	30,958 (67.9)	3,600 (40.9)	114,289 (81.0)	10,091 (76.6)
1	48,815 (18.9)	12,681 (27.3)	27,536 (26.2)	8,360 (31.4)	10,358 (22.7)	2,908 (33.0)	20,348 (14.4)	2,157 (16.4)
≥2	17,726 (6.9)	7,433 (16.0)	11,314 (10.8)	4,837 (18.2)	4,271 (9.4)	2,294 (26.1)	6,467 (4.6)	917 (7.0)
Exempt from copayment, n (%)*	182,941 (70.8)	37,567 (80.8)	75,418 (71.8)	21,334 (80.0)	36,812 (80.8)	7,747 (88.0)	98,629 (69.9)	10,411 (79.1)
Received geriatric care, n (%)	16,401 (6.3)	3,519 (7.6)	11,434 (10.9)	2,337 (8.8)	4,131 (9.1)	1,061 (12.0)	5,757 (4.1)	412 (3.1)
Number of primary care visits, n (%)								
0–1	42,860 (16.5)	4,813 (10.3)	21,597 (20.6)	3,088 (11.6)	4,797 (10.5)	712 (8.1)	20,217 (14.3)	1,134 (8.6)
2–4	131,751 (51.0)	20,948 (45.1)	52,691 (50.1)	12,236 (45.9)	19,372 (42.5)	3,198 (36.3)	72,114 (51.1)	6,215 (47.2)
≥5	83,949 (32.5)	20,720 (44.6)	30,791 (29.3)	11,326 (42.5)	21,418 (47.0)	4,892 (55.6)	48,773 (34.6)	5,816 (44.2)

All comparisons significant ($P < .01$) except comparisons of marital status for dementia and falls Rx-DIS and sex for chronic renal failure Rx-DIS.
* 0.02% not classified.
SD = standard deviation.

Rx-DIS included in the HEDIS measure than the more-comprehensive measure used in that study.[10]

The current study demonstrated the importance of examining exposure and risk factors according to the specific type of Rx-DIS exposure. Although rates of exposure for dementia and falls were between 15% and 20%, exposure for renal failure was only 8.5%, perhaps indicating more care when prescribing medications for patients with renal failure. Alternatively, the measure may underestimate exposure to NSAIDs that can be obtained over the counter because generic forms may be obtained outside the VA pharmacy system at relatively low cost.

Although some risk factors were consistently associated with Rx-DIS exposure (e.g., younger age, number of unique medications, psychiatric comorbidity, exempt from copayment), it was found that the relationship between certain characteristics varied according to the specific Rx-DIS (e.g., race, sex, geriatric care and primary care utilization). The only race category that had consistently higher rates of Rx-DIS exposure than whites was Hispanic. For African Americans, likelihood of Rx-DIS exposure was lower than that for whites only in those with dementia; there was no significant difference for falls and a slightly higher likelihood for renal failure. These findings are consistent with higher rates of PIPE in whites based on the Beers criteria[3,25] and a broad literature identifying racial disparities in care for patients ranging from cardiac disease, hypertension, epilepsy, and mental health conditions.[27–29] Alternatively, examination of bivariate data according to race indicated that Hispanics were more likely to have multiple mental health conditions and that African Americans were more likely than whites to have arthritis, which may be treated with potentially problematic medications. Although the interaction effect was not statistically significant, variation in comorbidity profiles may influence Rx-DIS exposure.

Studies of PIPE based on the Beers criteria consistently find that women are at greater risk for exposure,[4,30] but the current study found that women were only at higher risk for dementia Rx-DIS exposure. Thus, it appears that higher risk for women is not comprehensive but is instead specific to dementia or the types of drugs that are problematic for dementia (e.g., anticholinergics, tricyclic antidepressants).

Geriatric care the prior year was associated with greater risk for Rx-DIS exposure overall, which is inconsistent with prior studies examining exposure to Beers criteria drugs, which have found that geriatric care reduced the likelihood of exposure.[30,31] However, examination of Rx-DIS exposure according to condition revealed less likelihood of dementia and renal Rx-DIS exposure and greater likelihood of falls Rx-DIS exposure. Because fall assessment and prevention is an important component of geriatric care, it is possible that the finding results from better screening and documentation of falls in the electronic medical record by geriatricians. Thus, these data may reflect not only a selection bias, with the most complicated patients being seen by geriatricians, but also a detection bias, with patients receiving care from a geriatrician

Table 3. Predictors of Drug–Disease Interactions in Community-Dwelling Department of Veterans Affairs Patients According to Condition or Disease

Characteristic	Odds Ratio (95% Confidence Interval)			
	Any	Dementia	Falls	Chronic Renal Failure
Demographic				
Age	1.0 (0.99–1.0)	0.98 (0.98–0.99)	0.98 (0.98–0.99)	0.97 (0.97–0.98)
Race or ethnicity (vs white)				
African American	0.98 (0.95–1.01)	0.88 (0.83–0.92)	1.03 (0.94–1.12)	1.17 (1.11–1.23)
Hispanic	1.25 (1.20–1.31)	1.07 (1.01–1.13)	1.19 (1.07–1.31)	1.35 (1.24–1.46)
Other	0.94 (0.89–1.03)	0.93 (0.82–1.06)	0.98 (0.81–1.19)	1.05 (0.91–1.21)
Female (vs male)	1.40 (1.31–1.50)	1.43 (1.30–1.57)	1.12 (1.00–1.26)	1.08 (0.90–1.29)
Unmarried (vs married)	0.97 (0.95–0.99)	0.93 (0.90–0.96)	0.96 (0.92–1.01)	1.10 (1.05–1.14)
Health status				
Number of unique medications	1.09 (1.09–1.09)	1.13 (1.12–1.13)	1.09 (1.08–1.09)	1.04 (1.04–1.05)
Selim Physical	0.97 (0.97–0.98)	0.96 (0.96–0.97)	0.91 (0.90–0.92)	0.99 (0.98–1.00)
Selim Mental Health (vs 0)				
1	1.63 (1.59–1.67)	1.17 (1.13–1.21)	2.08 (1.96–2.20)	1.01 (0.06–1.06)
≥2	2.36 (1.28–2.43)	1.33 (1.27–1.39)	3.50 (3.28–3.74)	1.19 (1.10–1.28)
Access-to-care factors				
Not exempt from copayment (vs exempt)	0.78 (0.75–0.80)	0.85 (0.82–0.89)	0.75 (0.70–0.81)	0.76 (0.60–0.80)
Received geriatric care	1.01 (0.97–1.06)	0.68 (0.64–0.71	1.26 (1.16–1.36)	0.69 (0.62–0.77)
Number of primary care visits (vs 0–1)				
2–4	1.12 (1.08–1.17)	1.28 (1.22–1.34)	0.94 (0.86–1.03)	1.38 (1.29–1.49)
≥5	1.21 (1.16–1.26)	1.31 (1.24–1.38)	0.89 (0.80–0.98)	1.47 (1.36–1.59)

being more likely to have conditions such as falls documented in the electronic medical record.

These data have several limitations. First, they reflect only medications and healthcare utilization received in the VA. It is possible that other medications were ordered and dispensed outside the VA because the study period was after the initiation of Medicare Part D (January 1, 2006).[32] However, because patterns of Rx-DIS and prescribing were similar in 2004, 2005, and 2006, it is likely that the potential effect on the findings is limited. Second, in identifying the Rx-DIS conditions, access was available only to VA administrative data. Although Medicare data may improve ascertainment of the conditions of interest, 2 years of prior data was available to assess comorbid conditions, which is the recommended period to identify chronic disease states within VA data.[18] With regard to healthcare utilization, the criterion requiring at least one inpatient or outpatient visit per year may have excluded individuals who frequently received outside care, biasing the cohort to individuals who were sicker. Examination of Rx-DIS exposure using looser inclusion criteria, but all available data since 2003, resulted in similar rates of exposure and similar risk factors.

Moreover, this study assessed Rx-DIS exposure before the final approval of the HEDIS Rx-DIS measure. Thus, providers were not aware of this quality measure at the time of the study. Because this measure includes many Rx-DIS described previously in the Beers and McLeod criteria and because the selected Rx-DIS are a small, but clinically important, component of these measures, the Rx-DIS themselves or data supporting the Rx-DIS were available to clinicians for a number of years.[1,2,6,11,33] Although it is possible that publication of this measure in 2007 may have led to reductions in Rx-DIS because this measure may have had broader diffusion into clinical practice, data provided by the NCQA suggest that rates of exposure have sustained small increases (approximately 1% increase for each condition) over the past 3 years.[16] It is possible that increases are due to more attention to documenting conditions such as falls, but similar increases occurred for all three conditions, including chronic kidney disease, which is less commonly undercoded. This study provides a baseline to determine the extent to which diffusion of information included in this measure is reflected in VA clinical practice in the future, preferably in FY11 or FY12.

Finally, because falls are routinely undercoded, it is likely that the estimates of Rx-DIs are conservative because only veterans with severe falls tend to be identified as such using administrative data.

This study found a high rate of exposure to potentially harmful Rx-DIS for patients with dementia, falls, and chronic renal failure but generally lower rates than were found in Medicare data during the same time period.[16] Other studies have not examined the link between exposure and adverse outcomes using this less-comprehensive HEDIS measure of Rx-DIS. The current study suggests that research examining outcomes should examine outcomes and risk factors for individual conditions separately before combining them as a single measure of exposure. If links between Rx-DIS are demonstrated, the VA is uniquely positioned to use health information technology to reduce Rx-DIS exposure by implementing prescribing alerts within the electronic medication order process. As a leader in geriatric care, the VA is also in a position to test interven-

tions such as use of pharmacists or geriatricians within Patient Aligned Care Teams (patient-centered medical home) to improve the quality of care for older veterans and to conduct investigations on how shared decision-making between patient and provider may contribute to, or be used to reduce, Rx-DIS exposure.

ACKNOWLEDGMENTS

Conflict of Interest: This study was supported by VA Health Services Research Grant IIR-06–062. Dr. Hanlon is supported by National Institute of Aging Grants R01AG027017, P30AG024827, T32 AG021885, K07AG033174, and R01AG034056; National Institute of Mental Health Grant R34 MH082682; National Institute of Nursing Research Grant R01NR010135; and Agency for Healthcare Research and Quality Grant R01HS017695. Dr. Horton is a recipient of Geriatric Academic Career Award KO1 HP00114-02. We acknowledge the contributions of Jeffrey Tabares and Ranjani Davalath in assisting with manuscript preparation.

MJP and MH are employed by the South Texas Veterans Health Care System, DRB and MEA are employed by the Edith Nourse Rogers Memorial VA Hospital, and JTH is employed by the VA Pittsburgh Healthcare System. This study was funded by VA Health Services Research and Development Service (HSR&D) IIR 06-062.

Dr. Pugh has received research funding from VA HSR&D DHI 09-237 (PI); VA HSR&D IIR-06-062 PI, Epilepsy Foundation PI, VA HSR&D PPO 09-295 PI, VA HSR&D IIR 02-274 PI, Pugh as co-investigator: VA HSR&D IIR 08-274, VA HSR&D SDR-07-042, IIR-05-121, IAF-06-080, IIR-09-335, SHP 08-140, TRX 01-091 Department of Defense CDMRP 09090014, NIH R01-NR010828, Pugh Speaker Honoraria: 2009 Kelsey Seybold Research Foundation $400.

Dr. Berlowitz has received funding from VA HSR&D RRP 09-112 as PI, VA HSR&D IIR-06-062 as co-investigator.

Author Contributions: Pugh: concept and design of the study, acquisition of data, analysis and interpretation of data, and preparation of manuscript. Starner and Berlowitz: design of the study, interpretation of data, and preparation of manuscript. Amuan: acquisition and analysis of data and preparation of manuscript. Marcum and Horton: interpretation of data and preparation of manuscript. Hanlon: concept and design of the study, interpretation of data, and preparation of manuscript.

Sponsor's Role: The sponsor had no role or influence in matters relating to research design, methods, subject recruitment, data collections, analysis, or preparation of paper.

REFERENCES

1. Beers MH. Explicit criteria for determining potentially inappropriate medication use by the elderly. Arch Intern Med 1997;157:1531–1536.
2. Fick DM, Cooper JW, Wade WE et al. Updating the beers criteria for potentially inappropriate medication use in older adults: Results of a US consensus panel of experts. Arch Intern Med 2003;163:2716–2724.
3. Zhan C, Sangl J, Bierman AS et al. Potentially inappropriate medication use in the community-dwelling elderly: Findings from the 1996 Medical Expenditure Panel Survey. JAMA 2001;286:2823–2829.
4. Goulding MR. Inappropriate medication prescribing for elderly ambulatory care patients. Arch Intern Med 2004;164:305–312.
5. Pugh MJ, Fincke BG, Bierman A et al. Potentially inappropriate prescribing in elderly veterans: Are we using the wrong drug, wrong dose, or wrong duration? J Am Geriatr Soc 2005;53:1282–1289.
6. McLeod PJ, Huang AR, Tamblyn RM et al. Defining inappropriate practices in prescribing for elderly people: A national consensus panel. Can Med Assoc J 1997;156:385–391.
7. Lindblad CI, Hanlon JT, Gross CR et al. Clinically important drug-disease interactions and their prevalence in older adults. Clin Ther 2006;28:1133–1143.
8. Curtis LH, Ostbye T, Sendersky V et al. Inappropriate prescribing for elderly Americans in a large outpatient population. Arch Intern Med 2004;164: 1621–1625.
9. Simon SR, Chan KA, Soumerai SB et al. Potentially inappropriate medication use by elderly persons in U.S. Health Maintenance Organizations, 2000–2001. J Am Geriatr Soc 2005;53:227–232.
10. Zhan C, Correa-de-Araujo R, Bierman AS et al. Suboptimal prescribing in elderly outpatients: Potentially harmful drug-drug and drug-disease combinations. J Am Geriatr Soc 2005;53:262–267.
11. Lindblad CI, Artz MB, Pieper CF et al. Potential drug-disease interactions in frail, hospitalized elderly veterans. Ann Pharmacother 2005;39:412–417.
12. Hastings SN, Sloane RJ, Goldberg KC et al. The quality of pharmacotherapy in older veterans discharged from the emergency department or urgent care clinic. J Am Geriatr Soc 2007;55:1339–1348.
13. Hastings SN, Schmader KE, Sloane RJ et al. Quality of pharmacotherapy and outcomes for older veterans discharged from the emergency department. J Am Geriatr Soc 2008;56:875–880.
14. Chrischilles EA, VanGilder R, Wright K et al. Inappropriate medication use as a risk factor for self-reported adverse drug effects in older adults. J Am Geriatr Soc 2009;57:1000–1006.
15. Lund BC, Carnahan RM, Egge JA et al. Inappropriate prescribing predicts adverse drug events in older adults. Ann Pharmacother 2010;44:957–963.
16. The state of health care quality: Value, variation and vulnerable populations 2009. National Committee for Quality Assurance [on-line]. Available at http://www.ncqa.org/Portals/0/Newsroom/SOHC/SOHC_2009.pdf Accessed August 6, 2010.
17. HEDIS 2008 NDC Lists. National Committee on Quality Assurance [on-line]. Available at http://www.ncqa.org/tabid/598/Default.aspx Accessed February 4, 2011.
18. Borzecki AM, Wong AT, Hickey EC et al. Identifying hypertension-related comorbidities from administrative data: What's the optimal approach? Am J Med Qual 2004;19:201–206.
19. Selim AJ, Fincke G, Ren XS et al. Comorbidity assessments based on patient report: Results from the Veterans Health Study. J Ambulatory Care Manage 2004;27:281–295.
20. Selim AJ, Berlowitz DR, Fincke G et al. Risk-adjusted mortality rates as a potential outcome indicator for outpatient quality assessments. Med Care 2002;40:237–245.
21. Schneeweiss S, Seeger JD, Maclure M et al. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. Am J Epidemiol 2001;154:854–864.
22. Pugh MJ, Copeland LA, Zeber JE et al. The impact of epilepsy on health status among younger and older adults. Epilepsia 2005;46:1820–1827.
23. Pugh MJ, Rosen AK, Montez-Rath M et al. Potentially inappropriate prescribing for the elderly: Effects of geriatric care at the patient and health care system level. Med Care 2008;46:167–173.
24. Hanlon JT, Fillenbaum GG, Schmader KE et al. Inappropriate drug use among community dwelling elderly. Pharmacotherapy 2000;20:575–582.
25. Pugh MJ, Hanlon JT, Zeber JE et al. Assessing potentially inappropriate prescribing in the elderly Veterans Affairs population using the HEDIS 2006 quality measure. J Manage Care Pharm 2006;12:537–545.
26. Lindblad CI, Hanlon JT, Gross CR et al. Drug-Disease Interactions and Health Outcomes in Frail Elderly Veterans. Poster presented at the Annual Meeting of the American Geriatrics Society. 2006.
27. Kressin NR, Petersen LA. Racial differences in the use of invasive cardiovascular procedures: Review of the literature and prescription for future research. Ann Intern Med 2001;135:352–366.
28. Kales HC, Blow FC, Bingham CR et al. Race, psychiatric diagnosis, and mental health care utilization in older patients. Am J Geriatr Psychiatry 2000;8:301–309.
29. Szaflarski M, Szaflarski JP, Privitera MD et al. Racial/ethnic disparities in the treatment of epilepsy: What do we know? What do we need to know? Epilepsy Behav 2006;9:243–264.
30. Bierman AS, Pugh MJ, Dhalla I et al. Sex differences in inappropriate prescribing among elderly veterans. Am J Geriatr Pharmacother 2007;5: 147–161.
31. Schmader KE, Hanlon JT, Pieper CF et al. Effects of geriatric evaluation and management on adverse drug reactions and suboptimal prescribing in the frail elderly. Am J Med 2004;116:394–401.
32. Resources on the Medicare Prescription Drug Benefit. Kaiser Family Foundation [on-line]. Available at http://www.kff.org/medicare/rxdrugbenefit.cfm Accessed October 16, 1909.
33. Ray WA, Griffin MR, Schaffner W et al. Psychotropic drug use and the risk of hip fracture. N Engl J Med 1987;316:363–369.

HEALTH CARE REFORM

Effectiveness of Collaborative Care for Depression in Human Immunodeficiency Virus Clinics

Jeffrey M. Pyne, MD; John C. Fortney, PhD; Geoffrey M. Curran, PhD; Shanti Tripathi, MS; J. H. Atkinson, MD; Amy M. Kilbourne, PhD; Hildi J. Hagedorn, PhD; David Rimland, MD; Maria C. Rodriguez-Barradas, MD; Thomas Monson, MD; Kathryn A. Bottonari, PhD; Steven M. Asch, MD; Allen L. Gifford, MD

Background: Depression is common among persons with the human immunodeficiency virus (HIV) and is associated with unfavorable outcomes.

Methods: A single-blind randomized controlled effectiveness trial at 3 Veterans Affairs HIV clinics (HIV Translating Initiatives for Depression Into Effective Solutions [HITIDES]). The HITIDES intervention consisted of an offsite HIV depression care team (a registered nurse depression care manager, pharmacist, and psychiatrist) that delivered up to 12 months of collaborative care backed by a Web-based decision support system. Participants who completed the baseline telephone interview were 249 HIV-infected patients with depression, of whom 123 were randomized to the intervention and 126 to usual care. Participant interview data were collected at baseline and at the 6- and 12-month follow-up visits. The primary outcome was depression severity measured using the 20-item Hopkins Symptom Checklist (SCL-20) and reported as treatment response (≥50% decrease in SCL-20 item score), remission (mean SCL-20 item score, <0.5), and depression-free days. Secondary outcomes were health-related quality of life, health status, HIV symptom severity, and antidepressant or HIV medication regimen adherence.

Results: Intervention participants were more likely to report treatment response (33.3% vs 17.5%) (odds ratio, 2.50; 95% confidence interval [CI], 1.37-4.56) and remission (22.0% vs 11.9%) (2.25; 1.11-4.54) at 6 months but not 12 months. Intervention participants reported more depression-free days during the 12 months (β = 19.3; 95% CI, 10.9-27.6; $P < .001$). Significant intervention effects were observed for lowering HIV symptom severity at 6 months (β = –2.6; 95% CI, –3.5 to –1.8; $P < .001$) and 12 months (β = –0.82; –1.6 to –0.07; $P = .03$). Intervention effects were not significant for other secondary outcomes.

Conclusion: The HITIDES intervention improved depression and HIV symptom outcomes and may serve as a model for collaborative care interventions in HIV and other specialty physical health care settings where patients find their "medical home."

Trial Registration: clinicaltrials.gov Identifier: NCT00304915

Arch Intern Med. 2011;171(1):23-31

DEPRESSION IS ONE OF THE most common mental health disorders that afflict persons infected with the human immunodeficiency virus (HIV).[1-3] Despite available and efficacious treatments for depression, evidence suggests underdiagnosis and undertreatment of depression in routine HIV care.[4,5] Depression in persons with HIV may be associated with accelerated HIV disease progression,[6,7] decreased immune functioning,[7-9] nonadherence to HIV medication regimens,[10-12] and increased risk of mortality.[13-15] Depression is a modifiable risk factor, and effective treatment may improve HIV outcomes.[16-18]

In general adult primary care, collaborative care for depression is effective[19] and cost-effective.[20-22] Collaborative care models are based on the chronic care model,[23] facilitating collaboration between primary care and specialty mental health care providers to improve the quality of depression care and outcomes. Compared with referral specialty mental health care models, collaborative care allows patients to receive care in more accessible and less stigmatizing settings.

To our knowledge, no one has tested the collaborative care model for depression in a long-term specialty physical health care setting. This is an important gap because, for many patients with complex chronic illnesses, specialty physical health care clinics become their primary source of health care or "medical home."[24] For several reasons, the treatment of HIV and depression may be a potential model for testing the exportability of a collaborative care approach, first developed for primary care settings, to specialty-driven

Author Affiliations are listed at the end of this article.

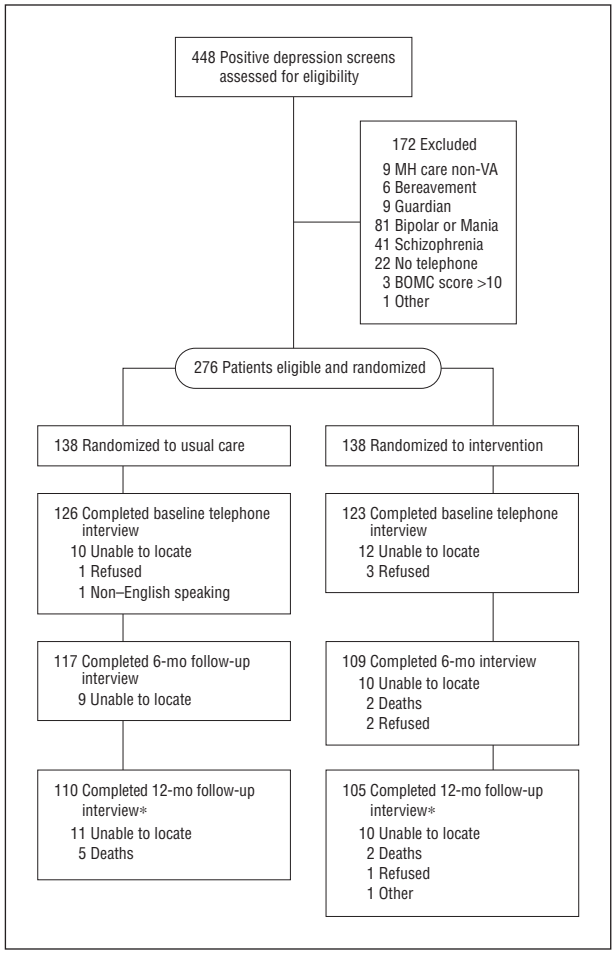


Figure 1. Flowchart of participants. *All participants who completed or were unable to be located for the 6-month interview were eligible for the 12-month interview. BOMC indicates Blessed Orientation-Memory-Concentration Test; MH, mental health; VA, Veterans Affairs.

long-term care of disease. Modern combination antiretroviral therapy has transformed HIV into a chronic disease, and the complexity of HIV care may be shifting care from primary to specialty care settings.[25] Organizational complexities are also associated with collaborative care interventions in HIV clinic settings given specialty care training, organizational structure, and culture.[26]

The goal of this study was to adapt an evidence-based primary care model of collaborative care of depression[27] to HIV clinic settings and evaluate the model's clinical effectiveness. We hypothesized that depressed patients who were assigned to the HIV Translating Initiatives for Depression Into Effective Solutions (HITIDES) intervention would report improved depression severity (primary outcome) and improved health-related quality of life, health status, HIV symptom severity, and medication regimen adherence (secondary outcomes) compared with patients receiving usual care.

METHODS

DESIGN

The HITIDES study was a randomized controlled effectiveness trial comparing depression collaborative care with enhanced usual care. Three Veterans Affairs (VA) HIV treatment facilities participated in the study. The study was approved by the research and development committees at the Central Arkansas Veterans Healthcare System, Michael E. DeBakey VA Medical Center, and Atlanta VA Medical Center.

PARTICIPANTS

Inclusion criteria were (1) a current 9-item Patient Health Questionnaire (PHQ-9) depression score of 10 or higher and (2) current treatment in the VA HIV clinic. A PHQ-9 score of at least 10 has strong psychometric properties in primary care settings (eg, >99% sensitivity and a 91% specificity).[28] Exclusion criteria were (1) no access to a telephone, (2) current acute suicidal ideation, (3) significant cognitive impairment as indicated by a score higher than 10 on the Blessed Orientation-Memory-Concentration Test,[29] and (4) history of bipolar disorder or schizophrenia. There were no exclusions based on physical health criteria, substance abuse or dependence, or current specialty mental health treatment.

The VA health care system provides health care to US veterans at more than 1400 medical centers and clinics nationwide. The VA is the largest national health care system and the largest single provider of HIV care. Veterans treated for HIV infection are older than the average US adult with HIV and more often male. Most of the 23 463 individuals in VA HIV care in 2008 were men (97%) and older than 50 years (64%). The proportion of African American patients (46%) is similar to the US HIV population (VA Clinical Case Registry, available at http://www.hiv.va.gov/). In general, veterans are eligible for care based on the length and character of their military service and current annual income level. In 2003, the VA stopped enrolling higher-income veterans.

After completing the eligibility and written informed consent processes, participants were randomly assigned to the intervention or to usual care and completed the baseline assessment. Participants were randomized to the intervention or to usual care in a 1:1 ratio according to a computer-generated random assignment sequence stratified by clinic and generated in advance. Research assistants at each clinic were provided envelopes labeled by participant number and containing randomized assignment. Participants were enrolled from February 1, 2007, through June 30, 2008.

DEPRESSION SCREENING AND INTERVENTION ADAPTATION BY CLINIC

Consistent with other collaborative care research, depression screening methods were adapted and adopted at each clinic as part of routine care using the PHQ-9 (2 clinics) or the 8-item PHQ (PHQ-8) (1 clinic)[30] and completed in person (2 clinics) or in person and by mail (1 clinic) before each HIV clinic visit. The intent was to complete the depression screen at each HIV clinic visit. The patient delivered a hard copy of the PHQ-9 or PHQ-8 to the HIV clinician at the time of the clinic visit. The PHQ-9 was completed by research personnel to confirm eligibility for the clinic that used the PHQ-8 as the depression screening instrument. A recent expert panel supported the use of the PHQ-9 in patients with HIV.[31]

Intervention adaptations were consistent across all clinics and included adding a clinical pharmacist to the intervention team, allowing substance-dependent patients to participate, adding brief alcohol and other drug interventions to the intervention, allowing patients engaged in specialty mental health care to participate, and formatting the intervention electronic medical record notes. The HIV staff conducted the depression screening, and the research team delivered the intervention.

HITIDES INTERVENTION

The purpose of the HITIDES intervention was to support HIV and mental health clinicians in delivering evidence-based depression treatment. The HIV depression care team consisted of a registered nurse depression care manager (DCM), a clinical pharmacist, and a psychiatrist (J.M.P.). This team was located offsite at the Little Rock VA Medical Center and convened once a week and as needed by telephone or in person. The depression care team communicated with treating clinicians via electronic medical record progress notes. The DCM communicated with patients via telephone. The HITIDES depression care team made treatment suggestions. Treatment decisions were made by the HIV or mental health clinicians at each site.

The DCM delivered the following intervention components: participant education and activation,[32] assessment of treatment barriers and possible resolutions, depression symptom and treatment monitoring, substance abuse monitoring, and instruction in self-management (eg, encouraging patients to exercise and participate in social activities).[33,34] The DCM used prewritten scripts, which are standardized instruments that were supported by the Web-based decision support system (NetDSS, available at https://www.netdss.net) during these telephone encounters.

The intervention used a stepped-care model for depression treatment.[35] The 5-step model included the following components plus DCM monitoring: (1) watchful waiting, (2) depression care team treatment suggestions (counseling or pharmacotherapy, considering participant preference), (3) pharmacotherapy suggestions after review of depression treatment history by the clinical pharmacist, (4) combination pharmacotherapy and specialty mental health counseling, and (5) referral to specialty mental health. Specific treatment suggestions were based on the Texas Medication Algorithm Project[36] and VA/Department of Defense Depression Treatment Guidelines.[37] Although the depression care team did not suggest watchful waiting, patient/provider treatment negotiations could result in this approach. At any time, HIV health care providers were free to refer participants directly to specialty mental health care. The stepped-care model was used to increase treatment intensity when participants did not respond to treatment.

The DCM conducted telephone-based monitoring every 2 weeks during acute treatment (before achieving a sustained 50% decrease in PHQ-9 score) and every 4 weeks during watchful waiting or continuation treatment (for 2 months after maintaining remission [PHQ-9 score, <5] or 6 months after maintaining a 50% decrease in the PHQ-9 score). The NetDSS system identified potential treatment nonresponse as (1) antidepressant regimen adherence of less than 80% during the past 14 days, (2) counseling nonadherence of less than 75% during the past month, (3) participant report of severe adverse effects during 2 consecutive DCM encounters, (4) participant report of a 5-point increase in depression severity from the enrollment PHQ-9 score based on 2 consecutive DCM encounters, or (5) lack of participant response (<50% decrease from enrollment PHQ-9 score during 2 consecutive DCM encounters) during an 8-week antidepressant or 12-week counseling trial.

USUAL CARE

Intervention and usual care participants completed PHQ depression screens as described, and patients delivered hardcopy results to their HIV clinicians at most clinic visits. These results were used to identify depression and monitor treatment response. Usual care depression treatment was provided by HIV or mental health clinicians without involvement of the HITIDES depression care team. Before starting the study, all HIV health care providers received 1 hour of HIV and depression training. Specialty mental health referral procedures were reviewed at all sites and typically included at least 1 failed depression treatment trial before referral.

DATA COLLECTION

Baseline and 6- and 12-month data were collected by telephone interviewers who were blinded to treatment assignment and used scripted computer-based assessments. At baseline, demographics (including self-reported race according to categories provided by the interviewers), depression history, and chronic physical health conditions (besides HIV) were measured using the Depression Outcomes Module.[38,39] Mental health comorbidity was measured using the Mini International Neuropsychiatric Interview.[40,41] Acceptability of antidepressant treatment was measured using an item developed for the Quality Improvement for Depression studies.[42,43] Follow-up data-collection interviews were completed for 226 of 249 participants (90.8%) at 6 months and 215 of 249 (86.3%) at 12 months (**Figure 1**).

OUTCOME MEASURES

The primary outcomes listed in clinicaltrials.gov were depression severity, implementation process, and quality of care. Implementation process and quality of care will be addressed in separate reports. Secondary outcomes were health status, health-related quality of life, HIV symptom severity, HIV medication regimen adherence, antidepressant regimen adherence, treatment satisfaction, and cost-effectiveness. Treatment satisfaction and cost-effectiveness also will be addressed in separate reports.

Depression symptom severity during the past 2 weeks was measured using the 20-item Hopkins Symptom Checklist (SCL-20),[44] which includes the 13-item depression scale plus 7 depression-related items from the Hopkins Symptom Checklist 90–Revised. The items are scored from 0 to 4 and averaged to provide a mean depression severity score ranging from 0 to 4.

Depression treatment response at 6 and 12 months was defined as a 50% or greater decrease in the mean SCL-20 score compared with baseline, and remission at 6 and 12 months was defined as a mean SCL-20 score of less than 0.5. Depression-free days (DFDs) were calculated as a summative measure of depression severity based on baseline and 6- and 12-month SCL-20 data using formulas originally developed by Lave and colleagues[45] and adapted for the SCL-20.[46] For each assessment, an SCL-20 score of 0.5 or less was considered depression free, a score of 2.0 or higher was considered fully symptomatic, and scores in between were assigned a linear proportional value. Derivation of DFDs from the SCL-20 has been used in several influential primary care studies.[47-49]

Health status was measured using the physical and mental health component summary scores from the Medical Outcomes Study Veterans 12-Item Short-Form Health Survey.[50] Health-related quality of life was measured using the Quality of Well-Being Self-administered Scale (QWB-SA).[51,52] The QWB-SA score is derived from general population preference weights and ranges from death (0.0) to perfect health (1.0). Severity of HIV symptoms was measured using the 20-item Symptoms Distress Module, which summarizes the degree to which each symptom bothered the participant in the past 4 weeks on a scale from 0 ("I do not have this symptom") to 4 ("This symptom bothers me a lot").[53] Bothersome HIV symptoms were defined as scores of 3 or 4.

Antidepressant and HIV medication regimen adherence were measured separately using the AIDS Clinical Trial Group assessment, which asks participants to report the number of pills per day they are supposed to take and the number of pills they

Table 1. Baseline Sociodemographic and Clinical Characteristics of Intervention and Usual Care Groups

Variable	Group[a]	
	Intervention (n=123)	Usual Care (n=126)
Sociodemographic		
Age, mean (SD), y	49.8 (8.7)	49.8 (10.5)
Male sex	120 (97.6)	122 (96.8)
African American race	78 (63.4)	77 (61.6)
Single/never married	103 (83.7)	98 (77.8)
High school graduate or higher	118 (95.9)	113 (89.7)
Annual income ≥$20 000	60 (50.8)	52 (42.6)
Clinical		
SF-12V PCS score, mean (SD)	41.5 (12.5)	39.5 (11.6)
SF-12V MCS score, mean (SD)	34.3 (10.5)	35.1 (11.0)
SCL-20 score, mean (SD)	1.8 (0.6)	1.9 (0.7)
QWB-SA score, mean (SD)[b]	0.49 (0.12)	0.44 (0.13)
Physical health comorbidity score, mean (SD)[c]	3.2 (2.3)	3.8 (2.3)
PHQ-9, mean (SD)[d]	15.7 (4.2)	16.0 (4.7)
Major depression[e]	92 (74.8)	98 (77.8)
Panic disorder[e]	10 (8.1)	18 (14.3)
Generalized anxiety disorder[e]	74 (60.2)	76 (60.3)
Posttraumatic stress disorder[e]	34 (27.6)	40 (31.7)
At-risk drinking[e]	19 (15.4)	26 (20.6)
Any inpatient mental health admission	33 (26.8)	32 (25.4)
Any past depression treatment	98 (79.7)	98 (77.8)
Any depression treatment in past 6 mo	68 (55.7)	67 (53.2)
Depression treatment type		
Watchful waiting acceptable	88 (71.5)	85 (67.5)
Antidepressant medication acceptable	88 (72.1)	87 (69.6)
Individual counseling acceptable	108 (87.8)	113 (89.7)
Group counseling acceptable	66 (53.5)	76 (60.3)
Bothersome HIV symptoms, mean (SD), No.	7.8 (4.1)	8.0 (4.3)
Current anti-HIV prescription	99 (80.5)	99 (78.6)
Skipped anti-HIV medication in past 4 d	23 (23.2)	28 (28.3)
Anti-HIV medication adherence, mean percentage (SD)	93.5 (16.2)	91.2 (20.1)
Current AD prescription	75 (61.0)	78 (61.9)
Skipped AD in past 4 d	22 (29.3)	20 (25.6)
AD regimen adherence, mean percentage (SD)	85.4 (30.5)	86.4 (31.1)

Abbreviations: AD, antidepressant; HIV, human immunodeficiency virus; MCS, mental component summary; PCS, physical component summary; PHQ-9, 9-item Patient Health Questionnaire; QWB-SA, Quality of Well-Being Self-administered Scale; SCL-20, 20-item Hopkins Symptom Checklist; SF-12V, Medical Outcomes Study Veterans 12-Item Short-Form Health Survey.

[a] Unless otherwise indicated, data are expressed as number (percentage) of participants. Percentages reflect the following missing data: race, 1 usual care participant; annual income, 5 intervention and 4 usual care participants; any depression treatment in the past 6 months, 1 intervention participant; and antidepressant acceptable, 1 intervention and 1 usual care participant.

[b] $P<.01$ for intervention vs usual care.

[c] $P<.05$ for intervention vs usual care.

[d] The PHQ-9 was used as depression screening measure. The SCL-20 was used as the depression outcome measure.

[e] Mental health comorbidity was identified using the Mini International Neuropsychiatric Interview.

skipped taking for each medication for each of the past 4 days.[54] Percentage of adherence equaled the total number of prescribed pills taken divided by the total number prescribed during the past 4 days. Dichotomous antidepressant regimen adherence was defined as greater than or equal to 80% adherence, and HIV medication regimen adherence was defined as greater than or equal to 95% adherence.

ANALYSIS

Participants were the unit of the intent-to-treat analysis. Sample size calculations were based on preliminary 6-month data from the Telemedicine Enhanced Antidepressant Management Study. The Telemedicine Enhanced Antidepressant Management Study was a depression collaborative care study conducted in VA community-based outpatient clinics. We based the power calculation on detecting an 11% difference in the percentage of responders between intervention and usual care using a 1-tailed t test ($\alpha=.05$). A sample size of 280 (140 subjects per arm) would provide 74% power. We did not adjust for potential nesting of participants within parent VA medical centers because the intraclass coefficient (0.02) was close to zero with respect to changes in SCL-20 scores $P=.30$), and there were no significant differences in outcomes across sites. Missing values were imputed using multiple imputation methods (ie, SAS statistical software, version 9.2, PROC MI and PROC MIANALYZE; SAS Institute Inc, Cary, North Carolina). Because of the large number of available case-mix variables, only those found to significantly predict dependent variables at $P<.20$ in bivariate analyses were included in multivariate analyses.

In **Table 1**, categorical variables were compared using a χ^2 test, and continuous variables were compared using a 2-tailed t test or its nonparametric analogue. Logistic and ordinary least squares regression analyses were used to estimate intervention effects for dichotomous and continuous outcomes, respec-

tively. For the continuous outcome measures DFDs and HIV symptom severity, the residual plots indicated nonconstant variance; therefore, we used weighted least squares regression methods to correct for heteroskedasticity. Residual plots for other continuous outcome measures did not indicate nonconstant variance. Separate regression analyses were conducted to examine the 6- and 12-month outcomes except for DFDs, which were measured for the entire 12-month follow-up period. Hierarchical linear modeling methods were not used because 2 of the primary outcomes (depression response and remission) were defined at only 2 time points, and for consistency we also used this approach with the secondary outcomes. The 9.2 version of the SAS software was used for all analyses.

RESULTS

In general, the HITIDES sample consisted of middle-aged, high school–educated, single, African American men of low to middle income who were experiencing mild to moderate symptoms of HIV disease and reported moderate reductions in health-related quality of life. Most had a history of mood disorder, expressed moderate depression symptom severity, and had received treatment for depression in the preceding 6 months. An important minority (18.1%) met criteria for at-risk drinking. Intervention and usual care participants were similar at baseline on demographic and clinical characteristics, except that intervention participants reported higher (better) health-related quality of life (QWB-SA scores, 0.49 vs 0.44; $P=.007$) and lower scores for physical health comorbidities (3.2 vs 3.8; $P=.046$) (Table 1).

INTERVENTION FIDELITY

Of the 123 intervention patients, 119 (96.7%) were contacted by the DCM. Initial patient education and activation was completed for 118 (99.2%), initial treatment barriers assessment was completed for 116 (97.5%), and 100% of all DCM contacts completed the PHQ-9 and medication regimen and/or counseling adherence assessment, depending on the current treatment. During the acute phase of treatment, a total of 231 intervention group treatment trials (mean, 1.94) included 110 (47.6%) watchful waiting, 94 (40.7%) pharmacotherapy, 7 (3.0%) counseling, and 20 (8.7%) combination pharmacotherapy and counseling trials. The mean number of DCM intervention telephone contacts per patient during the acute and continuation phases of treatment was 7.2 (SD, 4.5; range, 0-19).

PRIMARY OUTCOME: DEPRESSION SEVERITY

Figure 2 shows the unadjusted total SCL-20 scores over time. The unadjusted SCL-20 scores were not significantly different between the intervention and usual care groups at the 6- or 12-month follow-up. However, the unadjusted treatment response rates at 6 months were 17.5% (22 of 126 patients) for usual care and 33.3% (41 of 123 patients) for the intervention ($P=.004$) (**Table 2**). The adjusted intervention effect on treatment response at 6 months was also significant (odds ratio [OR], 2.60; 95% confidence interval [CI], 1.39-4.86; $P=.003$). The

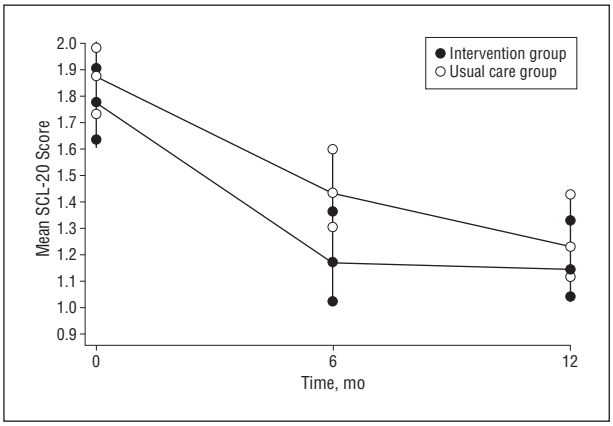


Figure 2. Unadjusted mean (95% confidence intervals) 20-item Hopkins Symptom Checklist (SCL-20) scores for the intervention and usual care groups over time.

unadjusted treatment remission rates at 6 months were 11.9% (15 of 126 patients) for usual care and 22.0% (27 of 123 patients) for the intervention ($P=.03$). The adjusted intervention effect on treatment remission at 6 months was also significant (OR, 2.40; 95% CI, 1.10-5.22; $P=.03$). Unadjusted and adjusted intervention effects on 12-month response and remission were not significant. Unadjusted (147.3 vs 120.0, $P=.04$) and adjusted intervention effects on DFDs were significantly greater in the intervention vs the control group ($\beta=19.3$; 95% CI, 10.9-27.6; $P<.001$) (**Table 3**).

SECONDARY OUTCOMES

Significant intervention effects were observed for HIV symptom severity but not for health-related quality of life, health status, antidepressant prescribing, or antidepressant or HIV medication regimen adherence. The adjusted intervention effect resulted in significantly lower 6- and 12-month HIV symptom severity compared with usual care at 6 months ($\beta=-2.6$; 95% CI, −3.5 to −1.8; $P<.001$) and 12 months ($\beta=-0.82$; 95% CI, −1.6 to −0.07; $P=.03$) (Table 3). The depression items in the SCL-20 and the HIV symptom severity measure overlapped (eg, baseline correlation, $r=0.54$; $P<.001$). However, after removing 7 depression-related items from the HIV symptom severity measure, the adjusted intervention effect on HIV symptom severity remained significant at 6 months ($\beta=-0.62$; 95% CI, −1.16 to −0.08; $P=.03$) but not at 12 months ($\beta=-0.09$; 95% CI, −1.58 to 1.40; $P=.88$). The unadjusted and adjusted intervention effects on 6- and 12-month antidepressant prescribing rates, antidepressant regimen adherence, HIV medication regimen adherence, QWB-SA, and Medical Outcomes Study Veterans 12-Item Short-Form Health Survey mental and physical component summary scores were not significant.

COMMENT

To our knowledge, this is the first effectiveness trial of a depression collaborative care intervention in a long-term specialty physical health care setting. Other successful collaborative depression care interventions tar-

Table 2. Unadjusted and Adjusted Dichotomous Outcome Results[a]

	Unadjusted			Adjusted		
	Group, No. (%) of Patients					
	Intervention	Usual Care	OR (95% CI)	OR (95% CI)	*P* Value	NNT[b]
Response[c]						
6 mo	41/123 (33.3)	22/126 (17.5)	2.50 (1.37-4.56)	2.60 (1.39-4.86)	.003	5
12 mo	49/123 (39.8)	41/126 (32.5)	1.37 (0.78-2.41)	1.29 (0.72-2.32)	.39	
Remission[d]						
6 mo	27/123 (22.0)	15/126 (11.9)	2.25 (1.11-4.54)	2.40 (1.10-5.22)	.03	7
12 mo	28/123 (22.8)	21/126 (16.7)	1.52 (0.78-2.98)	1.36 (0.66-2.88)	.40	
AD prescription[e]						
6 mo	72/108 (66.7)	78/115 (67.8)	0.89 (0.49-1.78)	0.89 (0.46-1.74)	.93	
12 mo	65/105 (61.9)	69/110 (62.7)	0.93 (0.49-1.78)	0.93 (0.49-1.78)	.98	
AD regimen adherence[f]						
6 mo	52/66 (78.8)	50/72 (69.4)	1.60 (0.74-3.45)	1.65 (0.75-3.62)	.22	
12 mo	45/59 (76.3)	51/60 (85.0)	0.55 (0.21-1.44)	0.56 (0.20-1.57)	.27	
HIV medication regimen adherence[g]						
6 mo	74/96 (77.1)	72/98 (73.5)	1.23 (0.63-2.40)	1.20 (0.60-2.31)	.65	
12 mo	68/92 (73.9)	64/86 (74.4)	0.93 (0.46-1.90)	1.60 (0.50-2.33)	.89	

Abbreviations: AD, antidepressant; CI, confidence interval; HIV, human immunodeficiency virus; NNT, number of patients needed to treat to achieve 1 additional successful outcome; OR, odds ratio.

[a] Attrition weights were used for AD prescription, AD regimen adherence, and HIV medication regimen adherence equations. Multiple imputation was used in other analyses.

[b] Number needed to treat was calculated for a given outcome only when the intervention *P* value was less than .05.

[c] The 6-month baseline covariates were intervention, 20-item Hopkins Symptom Checklist (SCL-20) score, physical health comorbidity, HIV symptom severity, race, any lifetime inpatient mental health admission, and HIV medication prescription. The 12-month baseline covariates were intervention, physical health comorbidity, HIV symptom severity, marital status, any lifetime inpatient mental health admission, any depression treatment in the past 6 months, and HIV medication prescription.

[d] The 6-month baseline covariates were intervention, SCL-20 score, physical health comorbidity, HIV symptom severity, race, comorbid mental health diagnosis, any lifetime inpatient mental health admission, any depression treatment in the past 6 months, and HIV medication prescription. The 12-month baseline covariates were intervention, age, SCL-20 score, physical health comorbidity, HIV symptom severity, and marital status.

[e] The 6-month baseline covariates were intervention, AD prescription, SCL-20 score, HIV symptom severity, race, annual household income, comorbid mental health diagnosis, any lifetime inpatient mental health admission, any depression treatment in the past 6 months, and acceptability of AD medications. The 12-month baseline covariates were intervention, AD prescription, SCL-20 score, physical health comorbidity, HIV symptom severity, race, comorbid mental health diagnosis, any lifetime inpatient mental health admission, any depression treatment in the past 6 months, and acceptability of AD medications.

[f] The 6-month baseline covariates were intervention, age, and education. The 12-month baseline covariates were intervention, age, education, and acceptability of AD medications.

[g] The 6-month baseline covariates were intervention, age, and race. The 12-month baseline covariates were intervention, age, HIV symptom severity, marital status, education, annual household income, any depression treatment in the past 6 months, and acceptability of AD medications.

geted depressed primary care patients generally,[48,55-58] depressed primary care patients with diabetes,[59] or depressed patients in an oncology clinic.[60] Depression is one of the most common co-occurring illnesses among people with complex chronic comorbidities; therefore, developing treatment strategies that are effective in specialty physical health care settings is vital to improving treatment outcomes for these patients. Human immunodeficiency virus may be a particularly important chronic condition model because depression is so prevalent and because treating depression can improve depression outcomes and has the potential to improve a wide range of life-saving self-management and adherence behaviors.[18,61]

The primary outcomes of this study were a more than doubling of the odds of depression response and remission at 6 but not 12 months and improved DFDs during the 12 months of treatment. Improved depression response and remission outcomes at 6 but not 12 months suggests that depression symptoms improved more rapidly in the intervention group compared with the usual care group. By 12 months, usual care participants caught up with intervention participants in terms of response and remission rates. The adjusted incremental 12-month DFDs result from the HITIDES intervention (19.3 days) compares with 20 to 72 DFDs in non-VA samples[45,56,62,63] and 14.6 incremental DFDs during 9 months in a VA sample.[22] Secondary outcomes included improved HIV symptom severity but no improvement in health-related quality of life, health status, or self-reported antidepressant or HIV medication regimen adherence.

The usual care depression response and remission outcomes appear to catch up with the intervention group at 12 months. Possible explanations include the following. First, the intervention was tested in settings where clinicians clearly accepted the need for improving depression recognition and treatment in the HIV clinic setting. Second, depression screening was completed on a hard copy form that most patients presented to their HIV clinician at every visit; therefore, over time, the HIV clinicians became more familiar with depression diagnosis and tracking treatment response. Third, DCM notes for intervention patients resulted in HIV clinicians becoming more familiar with treatment options for all patients in the HIV clinic.

Significant improvement in depression and HIV outcomes and the lack of detectable differences in prescribing or adherence suggest that other mechanisms lead to

Table 3. Unadjusted and Adjusted Continuous Outcome Results

	Unadjusted				Adjusted	
	Groups, Intervention Effects					
	Intervention	Usual Care	*P* Value	Effect Size[a]	Group Difference, β (95% CI)	*P* Value
DFDs[b]						
12 mo	147.3	120.0	.04	0.3	19.3 (10.9 to 27.6)	<.001
HIV severity[c]						
6 mo	−7.6	−4.5	.06	−0.2	−2.6 (−3.5 to −1.8)	<.001
12 mo	−7.9	−7.3	.75	−0.04	−0.8 (−1.6 to −0.07)	.03
QWB-SA[d]						
6 mo	0.02	0.005	.51		0.03 (−0.01 to 0.06)	.16
12 mo	0.01	0.04	.12		−0.01 (−0.05 to 0.03)	.49
SF-12V MCS score[e]						
6 mo	5.8	3.7	.26		2.0 (−1.0 to 5.0)	.19
12 mo	7.1	5.8	.50		1.7 (−1.7 to 5.2)	.32
SF-12V PCS score[f]						
6 mo	0.3	−0.1	.79		1.9 (−1.0 to 4.9)	.20
12 mo	1.7	0.9	.62		0.5 (−2.3 to 3.4)	.71

Abbreviations: CI, confidence interval; DFDs, depression-free days; HIV, human immunodeficiency virus; MCS, mental component summary; PCS, physical component summary; QWB-SA, Quality of Well-Being Self-administered Scale; SF-12V, Medical Outcomes Study Veterans 12-Item Short-Form Health Survey.

[a] Effect sizes were not calculated when the intervention effect was not significant in the adjusted analysis.

[b] Covariate baseline measures were intervention, 20-item Hopkins Symptom Checklist (SCL-20) score, physical health comorbidity, HIV symptom index, marital status, annual household income, comorbid mental health, any inpatient mental health admission, and any depression treatment in the past 6 months.

[c] The 6-month baseline covariates were intervention, HIV symptom severity, SCL-20 score, physical health comorbidity, marital status, education, comorbid mental health, any inpatient mental health admission, any depression treatment in the past 6 months, acceptability of antidepressant (AD) medications, and HIV medication prescription. The 12-month baseline covariates were intervention, HIV symptom severity, SCL-20 score, physical health comorbidity, marital status, annual household income, comorbid mental health, any inpatient mental health admission, any depression treatment in the past 6 months, and HIV medication prescription.

[d] The 6-month baseline covariates were intervention, QWB-SA score, SCL-20 score, physical health comorbidity, HIV symptom severity, education, annual household income, comorbid mental health, any inpatient mental health admission, acceptability of AD medications, and HIV medication prescription. The 12-month baseline covariates were intervention, QWB-SA score, SCL-20 score, physical health comorbidity, HIV symptom severity, education, annual household income, comorbid mental health, any inpatient mental health admission, acceptability of AD medications, and HIV medication prescription.

[e] The 6-month baseline covariates were intervention, SF-12V MCS score, SCL-20 score, physical health comorbidity, HIV symptom severity, education, annual household income, comorbid mental health, any inpatient mental health admission, any depression treatment in the past 6 months, acceptability of AD medications, and HIV medication prescription. The 12-month baseline covariates were intervention, SF-12V MCS score, age, SCL-20 score, HIV symptom severity, comorbid mental health, any inpatient mental health admission, any depression treatment in the past 6 months, and HIV medication prescription.

[f] The 6-month baseline covariates were intervention, SF-12V MCS score, age, SCL-20 score, physical health comorbidity, HIV symptom severity, annual household income, and HIV medication prescription. The 12-month baseline covariates were intervention, SF-12V PCS score, SCL-20 score, physical health comorbidity, HIV symptom severity, marital status, and education.

improved depression outcomes. We may have failed to detect medication prescription or adherence effects because of self-report measurement error.[64,65] The intervention may have led to greater dose intensification or treatment switching not detected by our measurement methods. Another mechanism for depression improvements may have been DCM promotion of self-management activities and/or brief interventions for alcohol and other drug abuse. Possible mechanisms for improved HIV symptom severity include depression intensifying patients' experience of physical symptoms[66] and/or DCM promotion of patient self-management activities.

The organization of health care into medical homes to provide integrated, accessible, and comprehensive services to patients with chronic and complicated medical needs has been widely promoted.[67] High-quality medical home services from specialists would require provision of specialty and primary care treatments, first-contact/comprehensive care responsibility, and patient affiliation with the clinic as the central hub of care.[24] Long-term care clinics for HIV infection and other complex conditions could, with adoption of appropriate care models, satisfy these criteria. There are also specialty clinics that provide comprehensive care for a shorter period that could satisfy most if not all medical home criteria for that period (eg, clinics that provide long-term treatment for hepatitis C virus infection).

The HITIDES study design had strengths and weaknesses. Strengths included adapting an evidence-based collaborative care intervention to HIV clinic settings, generalizability to the real-world patient population (eg, substance-dependent patients were not excluded), use of Web-based decision support to ensure fidelity to the intervention protocol, and use of an electronic medical record to facilitate communication between the HITIDES depression care team and the HIV and mental health care providers. Weaknesses included the potential lack of generalizability from VA to other treatment settings and the use of self-reported medication adherence data. However, a general VA outpatient depression collaborative care intervention found a significant correlation between self-report and administrative adherence data.[68]

In conclusion, the HITIDES intervention was successfully implemented in HIV clinic settings and improved depression and HIV symptom outcomes. The HITIDES intervention may serve as a model for collaborative care interventions in other specialty physical health care settings where patients find their medical home.

Accepted for Publication: March 23, 2010.
Author Affiliations: Center for Mental Healthcare and Outcomes Research, Central Arkansas Veterans Healthcare System, and Educational Center and Psychiatric Research Institute, University of Arkansas for Medical Sciences (Drs Pyne, Fortney, and Curran and Mr Tripathi), North Little Rock; South Central Mental Illness Research, Little Rock, Arkansas (Drs Pyne, Fortney, and Curran and Mr Tripathi); Psychiatry Service, Veterans Affairs (VA) San Diego Healthcare System and Department of Psychiatry, University of California, San Diego, School of Medicine (Dr Atkinson); VA Ann Arbor Health Services Research and Development Center of Excellence, National Serious Mental Illness Treatment Research and Evaluation Center, and Department of Psychiatry, University of Michigan, Ann Arbor (Dr Kilbourne); Minneapolis VA Medical Center and Department of Psychiatry, School of Medicine, University of Minnesota, Minneapolis (Dr Hagedorn); VA Medical Center and Department of Infectious Disease, Emory University School of Medicine, Atlanta, Georgia (Dr Rimland); Michael E. DeBakey VA Medical Center and Departments of Medicine and Infectious Disease, Baylor College of Medicine, Houston, Texas (Dr Rodriguez-Barradas); Department of Infectious Disease, Central Arkansas Veterans Healthcare System (Dr Monson); VA New England Healthcare System and Center for Healthcare Quality, Outcomes, and Economic Research, Bedford, Massachusetts (Drs Bottonari and Gifford); Department of VA Greater Los Angeles Healthcare System, Los Angeles, California (Dr Asch); and Departments of Public Health and Medicine, Boston University, Boston, Massachusetts (Dr Gifford).
Correspondence: Jeffrey M. Pyne, MD, Center for Mental Healthcare and Outcomes Research, Central Arkansas Veterans Healthcare System, 2200 Fort Roots Dr (152/NLR), North Little Rock, AR 72114 (jmpyne@uams.edu).
Author Contributions: *Study concept and design:* Pyne, Fortney, Curran, Atkinson, Kilbourne, Hagedorn, Asch, and Gifford. *Acquisition of data:* Pyne, Fortney, Curran, Rimland, Rodriguez-Barradas, and Monson. *Analysis and interpretation of data:* Pyne, Fortney, Curran, Tripathi, Kilbourne, Bottonari, and Asch. *Drafting of the manuscript:* Pyne, Atkinson, and Kilbourne. *Critical revision of the manuscript for important intellectual content*: Pyne, Fortney, Curran, Tripathi, Atkinson, Kilbourne, Hagedorn, Rimland, Rodriguez-Barradas, Monson, Bottonari, Asch, and Gifford. *Statistical analysis:* Fortney, Tripathi, and Bottonari. *Obtained funding:* Pyne, Fortney, Curran, Kilbourne, and Gifford. *Administrative, technical, and material support:* Pyne, Kilbourne, Rimland, Rodriguez-Barradas, Monson, Asch, and Gifford. *Study supervision:* Pyne, Asch, and Gifford.
Financial Disclosure: None reported.
Funding/Support: This study was supported by grant MNT 05-152 from the Department of Veterans Affairs Health Services Research and Development Service (Dr Pyne).
Disclaimer: The Department of Veterans Affairs had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or in the preparation, review, and approval of the manuscript.
Additional Contributions: We thank the veterans who participated in this study and the members of our dedicated study team for their efforts: Betty (Sunny) Ackerman, RN, was the depression care manager for the intervention; Regina Stanley, BA, was the project coordinator; Valorie Shue, BA, provided technical writing support; and Mitzi Mosier, provided administrative support.

REFERENCES

1. Rabkin JG. HIV and depression: 2008 review and update. *Curr HIV/AIDS Rep.* 2008;5(4):163-171.
2. Ciesla JA, Roberts JE. Meta-analysis of the relationship between HIV infection and risk for depressive disorders. *Am J Psychiatry.* 2001;158(5):725-730.
3. Kilbourne AM, Justice AC, Rabeneck L, Rodriguez-Barradas M, Weissman S, Team VP; VACS 3 Project Team. General medical and psychiatric comorbidity among HIV-infected veterans in the post-HAART era. *J Clin Epidemiol.* 2001;54(1)(suppl 1):S22-S28.
4. Sueoka K, Goulet JL, Fiellin DA, et al. Depression symptoms and treatment among HIV infected and uninfected veterans. *AIDS Behav.* 2010;14(2):272-279.
5. Asch SM, Kilbourne AM, Gifford AL, et al; HCSUS Consortium. Underdiagnosis of depression in HIV: who are we missing? *J Gen Intern Med.* 2003;18(6):450-460.
6. Lyketsos CG, Hoover DR, Guccione M, et al. Depressive symptoms as predictors of medical outcomes in HIV infection: Multicenter AIDS Cohort Study. *JAMA.* 1993;270(21):2563-2567.
7. Leserman J. HIV disease progression: depression, stress, and possible mechanisms. *Biol Psychiatry.* 2003;54(3):295-306.
8. Evans DL, Ten Have TR, Douglas SD, et al. Association of depression with viral load, CD8 T lymphocytes, and natural killer cells in women with HIV infection. *Am J Psychiatry.* 2002;159(10):1752-1759.
9. Leserman J, Petitto JM, Gu H, et al. Progression to AIDS, a clinical AIDS condition and mortality: psychosocial and physiological predictors. *Psychol Med.* 2002;32(6):1059-1073.
10. Singh N, Squier C, Sivek C, Wagener M, Nguyen MH, Yu VL. Determinants of compliance with antiretroviral therapy in patients with human immunodeficiency virus: prospective assessment with implications for enhancing compliance. *AIDS Care.* 1996;8(3):261-269.
11. Gordillo V, del Amo J, Soriano V, González-Lahoz J. Sociodemographic and psychological variables influencing adherence to antiretroviral therapy. *AIDS.* 1999;13(13):1763-1769.
12. Ickovics JR, Hamburger ME, Vlahov D, et al; HIV Epidemiology Research Study Group. Mortality, CD4 cell count decline, and depressive symptoms among HIV-seropositive women: longitudinal analysis from the HIV Epidemiology Research Study. *JAMA.* 2001;285(11):1466-1474.
13. Mayne TJ, Vittinghoff E, Chesney MA, Barrett DC, Coates TJ. Depressive affect and survival among gay and bisexual men infected with HIV. *Arch Intern Med.* 1996;156(19):2233-2238.
14. Moskowitz JT. Positive affect predicts lower risk of AIDS mortality. *Psychosom Med.* 2003;65(4):620-626.
15. Farinpour R, Miller EN, Satz P, et al. Psychosocial risk factors of HIV morbidity and mortality: findings from the Multicenter AIDS Cohort Study (MACS). *J Clin Exp Neuropsychol.* 2003;25(5):654-670.
16. Brown JL, Vanable PA. Cognitive-behavioral stress management interventions for persons living with HIV: a review and critique of the literature. *Ann Behav Med.* 2008;35(1):26-40.
17. Horberg MA, Silverberg MJ, Hurley LB, et al. Effects of depression and selective serotonin reuptake inhibitor use on adherence to highly active antiretroviral therapy and on clinical outcomes in HIV-infected patients. *J Acquir Immune Defic Syndr.* 2008;47(3):384-390.
18. Safren SA, O'Cleirigh C, Tan JY, et al. A randomized controlled trial of cognitive behavioral therapy for adherence and depression (CBT-AD) in HIV-infected individuals. *Health Psychol.* 2009;28(1):1-10.
19. Williams JW Jr, Gerrity M, Holsinger T, Dobscha S, Gaynes B, Dietrich A. Systematic review of multifaceted interventions to improve depression care. *Gen Hosp Psychiatry.* 2007;29(2):91-116.
20. Schoenbaum M, Unützer J, Sherbourne C, et al. Cost-effectiveness of practice-initiated quality improvement for depression: results of a randomized controlled trial. *JAMA.* 2001;286(11):1325-1330.
21. Pyne JM, Rost KM, Zhang M, Williams DK, Smith J, Fortney J. Cost-effectiveness of a primary care depression intervention. *J Gen Intern Med.* 2003;18(6):432-441.
22. Liu CF, Hedrick SC, Chaney EF, et al. Cost-effectiveness of collaborative care for

depression in a primary care veteran population. *Psychiatr Serv*. 2003;54(5):698-704.
23. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. *Milbank Q*. 1996;74(4):511-544.
24. Kirschner N, Barr MS. Specialists/subspecialists and the patient-centered medical home. *Chest*. 2010;137(1):200-204.
25. Phillips BR, Mole LA, Backus LI, Halloran JP, Chang SW. Caring for veterans with HIV disease: fiscal year 2002. http://www.publichealth.va.gov/docs/caring_2002.pdf. Accessed August 31, 2010.
26. Fultz SL, Goulet JL, Weissman S, et al. Differences between infectious diseases-certified physicians and general medicine–certified physicians in the level of comfort with providing primary care to patients. *Clin Infect Dis*. 2005;41(5):738-743.
27. Fortney JC, Pyne JM, Edlund MJ, et al. A randomized trial of telemedicine-based collaborative care for depression. *J Gen Intern Med*. 2007;22(8):1086-1093.
28. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann*. 2002;32(9):509-515.
29. Katzman R, Brown T, Fuld P, Peck A, Schechter R, Schimmel H. Validation of a short orientation-memory-concentration test of cognitive impairment. *Am J Psychiatry*. 1983;140(6):734-739.
30. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613.
31. Pyne JM, Asch SM, Lincourt K, et al. Quality indicators for depression care in HIV patients. *AIDS Care*. 2008;20(9):1075-1083.
32. Rost K, Nutting PA, Smith J, Werner JJ. Designing and implementing a primary care intervention trial to improve the quality and outcome of care for major depression. *Gen Hosp Psychiatry*. 2000;22(2):66-77.
33. Simon GE, VonKorff M, Rutter C, Wagner E. Randomised trial of monitoring, feedback, and management of care by telephone to improve treatment of depression in primary care. *BMJ*. 2000;320(7234):550-554.
34. Fortney JC, Pyne JM, Edlund MJ, Robinson DE, Mittal D, Henderson KL. Design and implementation of the telemedicine-enhanced antidepressant management study. *Gen Hosp Psychiatry*. 2006;28(1):18-26.
35. Katon W, Von Korff M, Lin E, et al. Stepped collaborative care for primary care patients with persistent symptoms of depression: a randomized trial. *Arch Gen Psychiatry*. 1999;56(12):1109-1115.
36. Suehs B, Argo TR, Bendele SD, Crismon ML, Trivedi MH, Kurian B. Texas Medication Algorithm Project Procedural Manual: major depressive disorder algorithms. http://www.dshs.state.tx.us/mhprograms/pdf/TIMA_MDD_Manual_080608.pdf. Accessed November 30, 2009.
37. MDD Working Group. VA/DoD Clinical Practice Guideline for Management of Major Depressive Disorder (MDD). Version 2.0. http://www.healthquality.va.gov/Major_Depressive_Disorder_MDD_Clinical_Practice_Guideline.asp. Accessed November 30, 2009.
38. Smith GR Jr, Burnam A, Burns BJ, Cleary P, Rost KM. Depression Outcomes Module (DOM). In: American Psychiatric Association, ed. *Handbook of Psychiatric Measures*. Washington, DC: American Psychiatric Association; 2000:213-215.
39. Rost K, Smith GR, Burnam MA, Burns BJ. Measuring the outcomes of care for mental health problems: the case of depressive disorders. *Med Care*. 1992;30(5)(suppl):MS266-MS273.
40. Lecrubier Y, Sheehan DV, Weiller E, et al. The Mini International Neuropsychiatric Interview (MINI): a short diagnostic structured interview: reliability and validity according to the CIDI. *Eur Psychiatry*. 1997;12(5):224-231.
41. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *Eur Psychiatry*. 1997;12(5):232-241.
42. Rost KM, Nutting P, Smith J, Werner J, Duan N. Improving depression outcomes in community primary care practice: a randomized trial of the quEST intervention: Quality Enhancement by Strategic Teaming. *J Gen Intern Med*. 2001;16(3):143-149.
43. Wells KB, Sherbourne C, Schoenbaum M, et al. Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *JAMA*. 2000;283(2):212-220.
44. Derogatis L. *SCL-90-R: Administration, Scoring, and Procedures Manual*. 3rd ed. Minneapolis, MN: National Computer Systems Inc; 1994.
45. Lave JR, Frank RG, Schulberg HC, Kamlet MS. Cost-effectiveness of treatments for major depression in primary care practice. *Arch Gen Psychiatry*. 1998;55(7):645-651.
46. Simon GE, Katon WJ, VonKorff M, et al. Cost-effectiveness of a collaborative care program for primary care patients with persistent depression. *Am J Psychiatry*. 2001;158(10):1638-1644.
47. Katon W, Robinson P, Von Korff M, et al. A multifaceted intervention to improve treatment of depression in primary care. *Arch Gen Psychiatry*. 1996;53(10):924-932.
48. Unützer J, Katon W, Callahan CM, et al; IMPACT Investigators Improving Mood-Promoting Access to Collaborative Treatment. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA*. 2002;288(22):2836-2845.
49. Simon GE, Ludman EJ, Rutter CM. Incremental benefit and cost of telephone care management and telephone psychotherapy for depression in primary care. *Arch Gen Psychiatry*. 2009;66(10):1081-1089.
50. Jones D, Kazis L, Lee A, et al. Health status assessments using the Veterans SF-12 and SF-36: methods for evaluating outcomes in the Veterans Health Administration. *J Ambul Care Manage*. 2001;24(3):68-86.
51. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. *Health Serv Res*. 1976;11(4):478-507.
52. Kaplan RM, Anderson JP. The general health policy model: an integrated approach. In: Spiker B, ed. *Quality of Life and Pharmacoeconomics in Clinical Trials*. 2nd ed. Philadelphia, PA: Lippincott-Raven Publishers; 1996:309-321.
53. Justice AC, Holmes W, Gifford AL, et al; Adult AIDS Clinical Trials Unit Outcomes Committee. Development and validation of a self-completed HIV symptom index. *J Clin Epidemiol*. 2001;54(suppl 1):S77-S90.
54. Chesney MA, Ickovics JR, Chambers DB, et al; Patient Care Committee & Adherence Working Group of the Outcomes Committee of the Adult AIDS Clinical Trials Group (AACTG). Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: the AACTG adherence instruments. *AIDS Care*. 2000;12(3):255-266.
55. Katon W, Russo J, Sherbourne C, et al. Incremental cost-effectiveness of a collaborative care intervention for panic disorder. *Psychol Med*. 2006;36(3):353-363.
56. Simon GE, Katon WJ, Lin EH, et al. Cost-effectiveness of systematic depression treatment among people with diabetes mellitus. *Arch Gen Psychiatry*. 2007;64(1):65-72.
57. Rollman BL, Belnap BH, LeMenager MS, et al. Telephone-delivered collaborative care for treating post-CABG depression: a randomized controlled trial. *JAMA*. 2009;302(19):2095-2103.
58. Dobscha SK, Corson K, Perrin NA, et al. Collaborative care for chronic pain in primary care: a cluster randomized trial. *JAMA*. 2009;301(12):1242-1252.
59. Katon WJ, Von Korff M, Lin EH, et al. The Pathways Study: a randomized trial of collaborative care in patients with diabetes and depression. *Arch Gen Psychiatry*. 2004;61(10):1042-1049.
60. Ell K, Xie B, Quon B, Quinn DI, Dwight-Johnson M, Lee PJ. Randomized controlled trial of collaborative care management of depression among low-income patients with cancer. *J Clin Oncol*. 2008;26(27):4488-4496.
61. Kumar V, Encinosa W. Effects of antidepressant treatment on antiretroviral regimen adherence among depressed HIV-infected patients. *Psychiatr Q*. 2009;80(3):131-141.
62. Simon GE, Manning WG, Katzelnick DJ, Pearson SD, Henk HJ, Helstad CS. Cost-effectiveness of systematic depression treatment for high utilizers of general medical care. *Arch Gen Psychiatry*. 2001;58(2):181-187.
63. Katon WJ, Schoenbaum M, Fan MY, et al. Cost-effectiveness of improving primary care treatment of late-life depression. *Arch Gen Psychiatry*. 2005;62(12):1313-1320.
64. Lu M, Safren SA, Skolnik PR, et al. Optimal recall period and response task for self-reported HIV medication adherence. *AIDS Behav*. 2008;12(1):86-94.
65. Garber MC, Nau DP, Erickson SR, Aikens JE, Lawrence JB. The concordance of self-report with other measures of medication adherence: a summary of the literature. *Med Care*. 2004;42(7):649-652.
66. Atkinson M, Zibin S, Chuang H. Characterizing quality of life among patients with chronic mental illness: a critical examination of the self-report methodology. *Am J Psychiatry*. 1997;154(1):99-105.
67. Carrier E, Gourevitch MN, Shah NR. Medical homes: challenges in translating theory into practice. *Med Care*. 2009;47(7):714-722.
68. Fortney JC, Pyne JM, Edlund MJ, et al. Reasons for antidepressant non-adherence among veterans treated in primary care clinics. *J Clin Psychiatry*. In press.

AIDS Care
Vol. 23, No. 11, November 2011, 1374–1381

Transactional sex risk and STI among HIV-infected female sex workers and HIV-infected male clients of FSWs in India

Anita Raj[a,b]*, N. Saggurti[c], Debbie M. Cheng[d], Anindita Dasgupta[a], Carly Bridden[b], Manojkumar Pradeshi[e] and J. H. Samet[a,b]

[a]*Boston University School of Public Health, Community Health Sciences, Boston, MA, USA;* [b]*Boston Medical Center and Boston University School of Medicine, Medicine, Boston, MA, USA;* [c]*Population Council, HIV and AIDS Program, 142 Golf Links, New Delhi, New Delhi 110003, India;* [d]*Boston University Public Health, Biostatistics, Boston, MA, USA;* [e]*Network of Maharashtra by People Living with HIV & AIDS (NMP+), Pune, India*

(*Received 11 November 2010; final version received 16 February 2011*)

To describe sex risk behaviors of HIV-infected female sex workers (FSWs) and HIV-infected male clients of FSWs, to evaluate associations between risky transactional sex and number of unprotected transactional sex episodes, and to assess the association between unprotected transactional sex and self-reported sexually transmitted infection (STI). Adult HIV-infected FSWs ($n = 211$) and HIV-infected male clients ($n = 205$) were surveyed in Mumbai about demographics, STI, and past 90-day and past year sex and substance use histories. Gender-stratified Poisson regression models were used to evaluate associations between four risky transactional sex behaviors (number of transactional sex partners; alcohol use before transactional sex; anal transactional sex; and transactional sex with a known HIV-infected partner) and number of unprotected transactional sex episodes; logistic regression was used to assess the association between unprotected transactional sex and self-reported STI. Twenty-nine percent of females and 7% of males reported any unprotected transactional sex episodes in the past 90 days. Thirty-nine percent of females and 12% of males reported past year STI. Among males, a greater number of transactional sex partners was associated with more unprotected transactional sex episodes (adjusted incidence rate ratio [IRR] = 8.2, 95% confidence interval [CI] = 1.8–38.4 highest vs. lowest tertile), and any unprotected transactional sex was associated with a higher odds of self-reported STI in the past year (adjusted odds ratio [AOR] = 5.6, 95% CI = 1.4–22.4). For women, risky transactional sex behaviors were not associated with condom non-use, and unprotected sex was negatively associated with STI (AOR = 0.4, 95% CI = 0.2–0.9). Reports of condom use during transactional sex were high for these samples. However, standard predictors of unprotected transactional sex (i.e., greater number of partners) and STI (i.e., unprotected sex) only held true for males. Further research is needed to guide an understanding of sex risk and STI among HIV-infected FSWs in India.

Keywords: female sex workers; male clients of female sex workers; HIV-infected; sex risk; India

Introduction

India has more than two million HIV-infected people, and female sex workers (FSWs) and male clients of FSWs maintain substantially higher HIV prevalence rates than that seen in the general population (National AIDS Control Organization [NACO], 2007; NACO, 2010). While the national HIV rate in India is estimated at 0.2–0.3% (NACO, 2007, NACO, 2010; Joint United Nations Programme on HIV/AIDS (UNAIDS), 2007, 2010), such rates among FSWs and their male clients in high HIV epidemic states such as Maharashtra are 18 and 12%, respectively (NACO, 2007). Ongoing sex risk among these populations poses a threat both to the transactional sex-involved and general population. To date, however, secondary prevention efforts in general and those specific to HIV-infected transactional sex-involved populations have received little attention from national prevention plans (NACO, 2010). More focus is needed toward such efforts and will benefit from understanding the sex risks occurring among these groups.

Within India, there is little research on sex practices among HIV-infected men and women of any population, but particularly those at risk of transmitting the virus via heterosexual sex. As of October 2010, we identified only two published studies examining sex risk behaviors among heterosexual HIV-infected adults in India (Chakrapani, Newman, Shunmugam, & Dubrow, 2010; Sri Krishnan et al., 2007), in addition to work from the current study (Samet et al., 2010). Overall, these studies indicate that although condom use is reported by the majority of HIV-infected adults, inconsistent and non-condom use is common, and barriers to condom use include both behavioral issues such as alcohol as well as attitudinal factors. More detailed consideration of transactional sex

*Corresponding author. Email: anitaraj@bu.edu

ISSN 0954-0121 print/ISSN 1360-0451 online

http://dx.doi.org/10.1080/09540121.2011.565034
http://www.tandfonline.com

behaviors, such as several partners or alcohol use before transactional sex, factors linked with increased risk for HIV/sexually transmitted infection (STI) among FSWs and male clients in India generally (NACO, 2010), was not identified. Nor was there consideration of differences in risk by gender, a likely scenario given the strong affect of gender role dynamics on condom use in India and elsewhere.

To build upon this nascent but growing literature on sex risk among HIV-infected adults in India, the following study examines sex behaviors among HIV-infected adults reporting past year transactional sex involvement. In addition, this study assesses whether risky transactional sex behaviors (e.g., alcohol use at transactional sex) are associated with the number of unprotected transactional sex episodes. Secondarily, this study examined the association between unprotected sex and STI in these populations. Findings can be used to inform the development of secondary prevention efforts in India and with HIV-infected individuals involved with transactional sex, globally.

Methods

The Transactional sex and Alcohol: Justification for a research initiative (TAJ) research team completed surveys on HIV-infected FSWs and HIV-infected male clients of FSWs in Mumbai, India ($n = 416$) from November 2008 to February 2009. Participants were recruited by way of service and health agencies focused on these populations in Mumbai, India; details on these agencies are available in previous publication from this study (Samet et al., 2010).

HIV-infected outreach workers at partnering agencies reviewed client lists and selected every fifth individual from the list to be approached and screened for study participation. Those contacted were asked to come to their respective recruitment sites for eligibility screening for the research study. A total of 326 women and 418 men were contacted for study recruitment, of which 246 (75%) women and 372 (89%) men came into their respective recruitment sites. Of these, 216 women and 210 men met the study's eligibility criteria: 18 years or older; HIV-infected; and reporting sex trade involvement in the past year (i.e., selling sex for women, purchasing sex for men) *and* penile-vaginal or anal sex in the past 30 days. HIV infection was confirmed by medical records brought by the participants. Of those eligible for the study, 5 women and 5 men were unwilling to participate and complete their interviews; they were excluded from the analyses, providing the final sample size of 211 female and 205 male participants.

Participants received a 45 minute interviewer-administered survey in Hindi assessing demographics, alcohol use, sex risk behaviors, and health status. Instruments were developed in English, translated into Hindi and then reviewed by a study investigator fluent in both languages. Discrepancies were resolved in consultation with the US investigators. Participants were given 100 rupees ($2.50) as compensation for their time in this study. Study procedures were reviewed and approved by the Institutional Review Boards (IRB) of Boston University Medical Campus, Network in Maharashtra by People Living with HIV (NMP+) and the Indian Council of Medical Research.

Measures

Demographic data were single item measures including age, level of education, income, religion, marital status, and number of children.

Sex risk behaviors were assessed for each of the following types of sex partners: spouses, transactional sex partners, and other sex partners. For each of these types of partners, participants were asked the number of these partners they had, as well as the number of vaginal or anal sex episodes and the frequency of unprotected sex episodes (indicated by 1 = always, 2 = usually, 3 = sometimes, 4 = rarely, and 5 = never) they had with these partners in the past year. They were also asked, again by partner type, whether any of the partners were HIV-infected and what their frequency of alcohol use before sex was with these partners (indicated by 1 = always, 2 = usually, 3 = sometimes, sometimes, 4 = rarely, and 5 = never). In addition, for each of these types of partners, participants were asked, using a past 90-day timeframe, the number of vaginal sex episodes, unprotected vaginal sex episodes, and occasions of alcohol use before sex; using this same time frame, they were then asked if they had engaged in anal sex with each type of partner and the frequency of condom use during anal sex, if the behavior was reported. All of these items were used to provide descriptive data on sex behaviors of the population; in addition, using data on transactional sex partners in the past 90 days, the following primary outcome variable was developed: number of unprotected transactional sex acts in the past 90 days.

To determine STI history, participants were asked whether they had contracted a sexually transmitted disease other than HIV, such as syphilis, gonorrhea, Chlamydia, trichomoniasis, or genital warts in the past year. Those indicating yes to any of these were classified as having an STI, which also served as one of our outcome variables. We were not able to

confirm STI history with medical record. We assessed date of HIV diagnosis to determine length of time since diagnosis.

Data analysis

All analyses for this study were stratified by gender. Poisson regression models were used to assess factors associated with number of unprotected transactional sex episodes in the past 90 days, as this outcome was not normally distributed. The Pearson Chi-square correction was used to account for overdispersion in the data. Four risky transactional sex behaviors were the main independent variables of interest: (1) number of transactional sex partners in the past year (categorized using tertiles; (2) any alcohol use before transactional sex in the past year; (3) any anal transactional sex in the past 90 days; and (4) any known HIV-infected transactional sex partners. The regression models controlled for the following potential demographic confounders: age, religion (Hindu vs. other), any formal education, and marital status (currently vs. formerly vs. never married). A multivariable regression model was fit including the four main independent variables and potential confounders. To avoid possible collinearity, pairwise Spearman correlations between the independent variables and covariates were assessed prior to regression modeling, and no pair of variables with correlation greater than 0.40 was included in the same model. The association between each independent variable and number of unsafe sex acts was quantified using an incidence rate ratio (IRR).

Multiple logistic regression models adjusting for potential confounders were used to assess the association between any unprotected transactional sex in the past 90 days and self-reported past year STI diagnosis. These analyses also controlled for age, religion, education, marital status, and the above described four risky transactional sex behaviors. Secondary analyses were also conducted evaluating any unprotected transactional sex in the past year (rather than past 90 days) as the main independent variable. As described above, to minimize the potential for collinearity, no pair of variables included in the same model was highly correlated. Analyses were performed using SAS software (version 9.1; SAS Institute, Cary, NC).

Results

Demographic characteristics of study participants

Female participants ranged in age from 22 to 49 years; women largely reported no history of formal education (78%) and were unmarried (91%) (Table 1). Male participants ranged in age from 20 to 49 years. Few men reported no formal education (11%), and 63% were unmarried.

Sex partners, unprotected sex and transactional sex behaviors

Among women, sex with spouses and other non-transactional sex partners in the past year was not common; however, for those reporting these behaviors ($n = 18$ and $n = 33$, respectively), the vast majority ($> 70\%$) reported unprotected sex in these contexts (Table 2). All women reported past year transactional

Table 1. Demographic characteristics of HIV-Infected FSWs ($n = 211$) and HIV-infected male clients of FSWs ($n = 205$) in Mumbai, India.

	Female sample		Male sample	
Characteristic	Mean	SD	Mean	SD
Age, years	30.6	5.0	32.8	5.5
Number of months since diagnosis	6.0	8.9	28.5	36.5
Income, rupees, past month	3107.9	1912.4	4878.9	2004.2
	n	Percentage (%)	*n*	Percentage (%)
Married				
Currently married	20	9.5	76	37.1
reviously married	105	49.8	41	20.0
Never married	86	40.8	88	42.9
Religion				
Hindu	164	77.7	157	76.6
Muslim	25	11.8	19	9.3
Other	22	10.4	29	14.1
No formal education	165	78.2	23	11.2

sex, as it was an entry criterion for the study, and almost all women (98.6%) reported transactional sex in the past 3 months; 90.5% of women reported unprotected transactional sex in the past year, whereas 28.6% reported this behavior in the past 3 months.

Sex with spouses in the past year appeared more common for men than women in this sample (reported by 39.0% of males vs. 8.5% of females), as did sex with non-transactional sex partners (reported by 22.8% of males vs. 15.6% of females) (Table 2). Among men reporting sex with spouses ($n = 80$), one-third reported unprotected sex with these wives; among men reporting sex with other non-transactional sex partners ($n = 47$), almost two-thirds reported engaging in unprotected sex in this context.

All participants reported transactional sex in the past year, as it was an inclusion criterion for the study; as expected the mean number of partners reported by women in the past year was far greater than those reported by men: 554 partners (standard deviation [SD] = 28.7) vs. 9 partners (SD = 8.9), respectively (Table 2). As almost half of women reported HIV diagnosis in the past 3 months, we explored whether in the past 90-day unprotected sex was significantly different for those diagnosed in the past 3 months relative to those diagnosed more than 3 months ago; no significant association was observed. Among men, 26% reported unprotected transactional sex in the past year, and 7% reported unprotected transactional sex in the past 90 days (Table 2).

Alcohol use before transactional sex in the past year was common for both female and male participants, reported by 48 and 67%, respectively; 12% of females and 33% of males reported always using alcohol before transactional sex involvement (Table 3). Anal transactional sex in the past 90 days was reported by 5% of females but only 0.5% of males. Less than 4% of males and females reported transactional sex with known HIV-infected partners.

Past year STI and time since HIV diagnosis

Among females, 39% reported an STI in the past year (Table 3). Mean number of months since HIV diagnosis among women was 6 months (SD = 8.9 months). Among males, 12% reported STI in the past year. Mean number of months since diagnosis was 28.5 months for males (SD = 36.5 months).

Associations between risky transactional sex behaviors and unprotected transactional sex

Poisson analyses revealed no significant associations between higher risk transactional sex behaviors and number of unprotected transactional sex episodes

Table 2. Types of sex partners and unprotected sex practices by partner type among HIV-infected FSWs ($n = 211$) and HIV-infected clients of FSWs ($n = 205$) in Mumbai, India.

	Female sample				Male sample			
Characteristics	Median	IQR	*n*	Percentage (%)	Median	IQR	*n*	Percentage (%)
Sex with a spouse, past year			18	8.5			80	39.0
Number of sex episodes	120.0	(80.0–200.0)			70.0	(41.5–96.0)		
Any unprotected sex			13	72.2% (no./*n* = 13/18)			31	38.8% (no./*n* = 31/80)
Sex with other (non-transactional) partners, past year			33	15.6%			47	22.8%
Number of sex episodes	82.0	(40.0–300.0)			16.0	(7–50)		
Any unprotected sex			31	93.9% (no./*n* = 31/33)			30	63.8 (no./*n* = 30/47)
Transactional sex, past year			211	100%			205	100%
Number of sex episodes	650	(420.0–800.0)			6	(4–12)		
Number of partners	600	(360.0–720.0)			6	(4–10)		
Any unprotected transactional sex			191	90.5% (no./*n* = 191/211)			53	25.9 (no./*n* = 53/205)
Transactional sex, past 3 months			208	98.6%			199	97.1%
Number of sex episodes	90.0	(60.0–190.0)			2.0	(1–3)		
Any unprotected sex			60	28.6% (no./*n* = 60/211)			13	6.5% (no./*n* = 13/205)

Table 3. High risk transactional sex behaviors and STI among HIV-Infected FSWs ($n = 211$) and HIV-infected clients of FSWs ($n = 205$) in Mumbai, India.

	Female sample		Male sample	
	n	Percentage (%)	*n*	Percentage (%)
Frequency of alcohol use before transactional sex				
Always	25	11.8	68	33.3
Usually	11	5.2	13	6.4
Sometimes	39	18.6	54	26.5
Rarely	13	6.2	2	1.0
Never	122	57.8	67	32.8
Any anal transactional sex, past 3 months	11	5.2	1	.5
Any transactional sex with known HIV+ partners	8	3.8	7	3.4
Any STI, past year	82	38.9	25	12.3

among female participants; among males, only a high number of transactional sex partners was significantly associated with unprotected sex in the past 90 days (adjusted IRR = 8.2, 95% confidence interval [CI] = 1.8–38.4, highest vs. lowest tertile) (Table 4). Due to the large proportion of female participants reporting recent HIV diagnosis, exploratory analyses were conducted, stratifying the female sample by whether or not HIV diagnosis occurred in the past 3 months; no difference in findings was observed across these groups in terms of the association between higher risk transactional sex behaviors and number of unprotected transactional sex episodes.

Associations between unprotected transactional sex and STI

For women, unprotected sex in the past 90 days was associated with a lower odds of STI (AOR = 0.4, 95% CI = 0.2, 0.9), but for men, it was a significant risk factor for STI (AOR = 5.6, 95% CI = 1.4, 22.4) (Table 5). Again, exploratory analyses were conducted, stratifying the female sample by whether or not HIV diagnosis occurred in the past 3 months; no difference in findings was observed.

Among women, alcohol before transactional sex (AOR = 2.0, 95% CI = 1.0, 3.8) and having the highest number of transactional sex partners

Table 4. Associations between risky transactional sex behaviors and number of unprotected sex episodes in the past 90 days among HIV-Infected FSWs ($n = 211$) and HIV-infected male clients of FSWs ($n = 205$).

	Female sample		Male sample	
	Crude IRR (95% CI)	Adjusted IRR[a] (95% CI)	Crude IRR (95% CI)	Adjusted IRR[a] (95% CI)
Number of transactional sex partners, past year[b]				
Low	1.0	1.0	1.0	1.0
Mid-range	1.74 (0.73, 4.17)	1.69 (0.71, 3.98)	0.23 (0.01, 6.09)	0.28 (0.01, 6.37)
High	1.87 (0.79, 4.43)	1.75 (0.72, 4.23)	5.17 (1.06, 25.34)	8.21 (1.76, 38.39)
Any alcohol use before transactional sex, past year	1.14 (0.61, 2.15)	1.07 (0.55, 2.09)	1.72 (0.32, 9.26)	1.32 (0.34, 5.06)
Any anal transactional sex, past 90 days	1.17 (0.32, 4.32)	1.17 (0.31, 4.47)	*	*
Any transactional sex with HIV+ partner, past 90 days	1.80 (0.51, 6.36)	1.49 (0.41, 5.45)	*	*
Any formal education	1.27 (0.63, 2.57)	1.05 (0.51, 2.16)	1.00 (0.10, 10.34)	0.62 (0.09, 4.26)
Muslim/other religion vs. Hindu	1.45 (0.73, 2.88)	1.67 (0.81, 3.43)	0.56 (0.07, 4.12)	0.60 (0.11, 3.33)
Age $\leq$ 30 years	1.84 (0.91, 3.72)	1.90 (0.89, 4.06)	0.07 (0.01, 0.68)	0.09 (0.01, 0.81)
Marital status				
Former vs. current	0.83 (0.28, 2.49)	1.15 (0.38, 3.50)	0.27 (0.04, 1.96)	0.22 (0.04, 1.29)
Never vs. current	0.87 (0.29, 2.64)	0.97 (0.32, 2.97)	0.21 (0.04, 1.02)	0.27 (0.06, 1.24)

[a]Poisson regression models with overdispersion adjusted for age, any formal education, marital status, and religion.
[b]Categorized based on tertiles. For women, low = 10–440 partners, mid-range = 450–672 partners, highest = 799–1344 partners. For men, low = 1–4 partners, mid-range = 5–8 partners, high = 9–80 partners.
*Reported by too few participants to allow for analyses.

Table 5. Association between unprotected transactional sex and past year self-reported STI among HIV-infected FSWs ($n = 211$) and HIV-infected male clients of FSWs ($n = 205$) in Mumbai, India.

	Female sample		Male sample	
	Crude OR (95% CI)	AOR[a] (95% CI)	Crude OR (95% CI)	AOR[a] (95% CI)
Any unprotected transactional sex, past 90 days	0.55 (0.29, 1.04)	0.44 (0.22, 0.88)	4.91 (1.47, 16.40)	5.62 (1.41, 22.43)
Any alcohol use at transactional sex	1.59 (0.91, 2.78)	1.97 (1.04, 3.75)	1.30 (0.55, 3.47)	1.30 (0.48, 3.50)
Any anal transactional sex	0.33 (0.07, 1.57)	0.25 (0.05, 1.29)	*	*
Any HIV+ transactional sex partners	0.92 (0.21, 3.96)	0.92 (0.19, 4.55)	*	*
No. of transactional sex partners (Tertiles)[b]				
Lowest	1.0	1.0	1.0	1.0
Mid	0.88 (0.43, 1.79)	1.01 (0.47, 2.15)	0.63 (0.23, 1.72)	0.91 (0.29, 2.91)
Highest	2.13 (1.07, 4.22)	3.01 (1.35, 6.72)	0.68 (0.26, 1.81)	1.04 (0.34, 3.19)
Any formal education	1.02 (0.52, 1.98)	0.86 (0.41, 1.82)	0.88 (0.24, 3.21)	1.20 (0.28, 5.06)
Muslim/other religion vs. Hindu	0.97 (0.50, 1.89)	1.36 (0.63, 2.93)	0.81 (0.29, 2.29)	0.60 (0.19, 1.87)
Age ≤ 30 years	1.30 (0.73, 2.32)	1.22 (0.63, 2.35)	0.69 (0.30, 1.60)	0.60 (0.21, 1.71)
Marital status				
Former vs. current	2.25 (0.76, 6.65)	3.60 (1.04, 12.44)	1.64 (0.47, 5.76)	2.13 (0.55, 8.32)
Never vs. current	1.78 (0.59, 5.35)	1.95 (0.57, 6.63)	2.18 (0.79, 5.98)	3.50 (1.00, 12.29)

[a]Logistic regression models adjusted for age, any formal education, marital status, religion, and all other risky transactional sex behaviors.
[b]This variable was created based on tertiles. For women, low = 10–440 partners, mid-range = 450–672 partners, highest = 799–1344 partners. For men, low = 1–4 partners, mid-range = 5–8 partners, high = 9–80 partners.
*Reported by too few participants to allow for analyses.

(AOR = 3.0, 95% CI = 1.4, 6.7, highest vs. lowest tertile) were, however, significantly associated with STI; though this was not revealed to be the case for men. In addition to these behavioral predictors of STI, marital status was also found to be significantly associated with STI for both females and males. For females, those formerly married had a higher odds of reporting an STI diagnosis in the past year compared with those currently married (AOR = 3.6, 95% CI = 1.0, 12.4). For males, those never married had significantly higher odds of reporting an STI in the past year compared to those currently married (AOR = 3.5, 95% CI = 1.0, 12.3).

Discussion

The current study documents high rates of condom use among HIV-infected men and women involved in transactional sex, a finding consistent with that seen in previous studies with non-transactional sex involved HIV-infected adults in India (Chakrapani et al., 2010; Sri Krishnan et al., 2007). Such findings are also consistent with growing evidence from India indicating a slowing of the HIV epidemic and substantial reductions in HIV transmission among those involved in transactional sex, which have largely been attributable to increased condom use in this population (NACO, 2007, NACO, 2010; UNAIDS, 2007, 2010).

Nonetheless, unprotected sex is still occurring among a substantial number of these individuals, with patterns differing for women and men. Although the proportion of sex episodes reported as unprotected is relatively low for both male and female participants, it was more likely among FSWs. Further, the large number of sex episodes reported by HIV-infected FSWs (a median of 90 episodes in the past 3 months) maintains a notable number of unprotected transactional sex encounters in this population. Male clients, in contrast, were more likely than FSWs to report non-transactional sex partners, with unprotected sex being more likely with these partners than with FSWs. Overall, these findings indicate that the risk for sex transmission among this HIV-infected sample persists, though to different classes of partners for FSWs and male clients.

As unprotected sex patterns varied by gender, so did associations between unprotected transactional sex and past year STI diagnosis. For men, a greater number of transactional sex partners was associated with more unprotected transactional sex, which in turn was associated with a higher odds of past year STI diagnosis. These findings clearly support the potential utility of behavioral risk reduction approaches to intervention for this population. For women, we were unable to detect associations between risky transactional sex behaviors and frequency

of unprotected sex, and unprotected sex was actually associated with lower odds of past year STI. These conflicting findings for females may be indicative of their having less control than male partners or others over condom use in the transactional sex encounter. The observed relationship between unprotected sex being associated with STI may be a consequence of the timeframes used to assess this association or the cross-sectional nature of the study. Such findings demonstrate the need for greater research to understand unprotected sex and STI risk among HIV-infected sex workers, with an eye toward what interventions can reduce these risks.

Some insight into this issue is provided via our secondary analyses, which demonstrate higher odds of STI among FSWs reporting the highest number of transactional sex partners and alcohol use before the transactional sex encounter. Increased STI risk as a consequence of greater exposure from more partners is understandable, but increased risk for STI among those FSWs using alcohol is less clear, particularly as current findings did not demonstrate a significant association between alcohol use and the frequency of unprotected transactional sex episodes. Rather, use of alcohol before transactional sex may be a marker for a context of increased risk. In India, there are venues that provide alcohol and sex together for clients (Go et al., 2011; Schensul, Singh, Gupta, Bryant, & Verma, 2010), and these may pose a greater risk for STI among FSWs working within them. More research is needed to examine whether this in fact may be the case and the potential utility of alcohol venue-based interventions to address both HIV and STI transmission between FSWs and male clients.

Limitations

The current study must be considered in light of a number of study limitations. The sample was drawn from individuals linked to agencies serving HIV-infected FSWs and men in a single metropolitan area (Mumbai), limiting generalizability of findings. In addition, much of the data in this study came from self-report and is thus subject to both social desirability and recall biases. Finally, associations are based on the use of different timeframes for the assessment period for certain variables and cross-sectional rather than longitudinal data; the latter precludes assumptions of causality based on the observed associations. More research with larger samples and longitudinal data would be useful to confirm current findings and explore in greater detail issues unable to be examined in this study due to reported low frequency events, such as anal sex and sex with HIV-infected partners. The latter issues have been seen to be a major predictor of STI in a previous study within India (Subramanian et al., 2008).

Conclusion

The current study contributes to the small but growing literature on sex risk among HIV-infected adults in India by examining these issues among those at greatest risk for transmission – HIV-infected FSWs and HIV-infected male clients of FSWs. The findings document that ongoing unprotected sex persists at low levels, concentrated for FSWs within transactional sex, and for male clients within non-transactional sex (e.g., sex with wives). In addition, among women, issues of alcohol use at transactional sex and very high numbers of transactional sex partners heighten their odds for STI, whereas unprotected sex did not appear to do this; in contrast among men, unprotected sex does heighten the risk for STI. These findings support the need for gender-tailored secondary HIV prevention efforts in India.

Acknowledgements

Funded by the National Institute on Alcohol Abuse and Alcoholism (R01AA016059-S1 and K24AA015674) and the Indian Council of Medical Research. We thank Ms. Shabana Patel from the NMP+ for her oversight of study implementation in terms of helping ensure safe, confidential, and respectful treatment of our HIV-infected participants. We also thank the members of our partnering community based organizations in Mumbai for their assistance in implementing this study in the field. From the NPM+ in Mumbai, this includes Ms. Bharti Sonawane, Ms. Aruna Bhere, Ms. Sunita, Mr. Anil Dalvi, and Mr. Hement Pawar. From ASHA Mahila Sanshta, this includes Ms. Mailee Tamang, Ms. Rupa, and Mr. Santosh Patankar. Finally, we are immensely grateful to the participants of this study who gave their time and shared their personal information and experiences in relationships and sex encounters.

References

Chakrapani, V., Newman, P.A., Shunmugam, M., & Dubrow, R. (2010). Prevalence and contexts of inconsistent condom use among heterosexual men and women living with HIV in India: Implications for research. *AIDS Patient Care and STDs*, *24*(1), 49–58.

Go, V.F., Srikrishnan, A.K., Parker, C.B., Salter, M., Green, A.M., Sivaram, S., ... Celentano, D.D. (2011). High prevalence of forced sex among non-brothel based, wine shop centered sex workers in Chennai, India. *AIDS and Behavior*, *15*(1), 163–171.

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2007). AIDS epidemic update: December 2007.

Retrieved from http://data.unaids.org/pub/epislides/2007/2007_epiupdate_en.pdf

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2010). UNAIDS report on the global AIDS epidemic. Retrieved from http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2010/20101123_globalreport_en.pdf

National AIDS Control Organization. (2007). HIV sentinel surveillance and HIV estimation in India 2007: A technical brief. Retrieved from http://www.nacoonline.org/upload/Publication/M&E%20Surveillance,%20Research/HIV%20Sentinel%20Surveillance%20and%20HIV%20Estimation%202007_A%20Technical%20Brief.pdf

National AIDS Control Organization. (2010). Annual report, 2009–2010. Retrieved from http://nacoonline.org/upload/AR%202009-10/NACO_AR_English%20corrected.pdf

Samet, J.H., Pace, C.A., Cheng, D.M., Coleman, S., Bridden, C., Pardesi, M., … Raj, A. (2010). Alcohol use and sex risk behaviors among HIV-infected female sex workers (FSWs) and HIV-infected male clients of FSWs in India. *AIDS and Behavior*, *14*(Suppl. 1), 74–83.

Schensul, J.J., Singh, S.K., Gupta, K., Bryant, K., & Verma, R. (2010). Alcohol and HIV in India: A review of current research and intervention. *AIDS and Behavior*, *14*(Suppl. 1), 1–7.

Sri Krishnan, A.K., Hendriksen, E., Vallabhaneni, S., Johnson, S.L., Raminani, S., Kumarasamy, N., … Safren, S.A. (2007). Sex behaviors of individuals with HIV living in South India: A qualitative study. *AIDS Education & Prevention*, *19*(4), 334–345.

Subramanian, T., Gupte, M.D., Paranjape, R.S., Brahmam, G.N., Ramakrishnan, L., Adhikary, R., … IBBA Study Team. (2008). HIV, sexually transmitted infections and sexual behaviour of male clients of female sex workers in Andhra Pradesh, Tamil Nadu and Maharashtra, India: Results of a cross-sectional survey. *AIDS*, *22* (Suppl. 5), 69–79.

Effects of Dram Shop Liability and Enhanced Overservice Law Enforcement Initiatives on Excessive Alcohol Consumption and Related Harms

Two Community Guide Systematic Reviews

Veda Rammohan, MPH, Robert A. Hahn, PhD, MPH, Randy Elder, PhD, Robert Brewer, MD, MSPH, Jonathan Fielding, MD, MPH, MBA, Timothy S. Naimi, MD, MPH, Traci L. Toomey, PhD, Sajal K. Chattopadhyay, PhD, Carlos Zometa, PhD, MSPH, Task Force on Community Preventive Services

Context: Dram shop liability holds the owner or server(s) at a bar, restaurant, or other location where a patron, adult or underage, consumed his or her last alcoholic beverage responsible for harms subsequently inflicted by the patron on others. Liability in a state can be established by case law or statute. Overservice laws prohibit the sale of alcoholic beverages to intoxicated patrons drinking in on-premises retail alcohol outlets (i.e., premises where the alcohol is consumed where purchased); enhanced enforcement of these laws is intended to ensure compliance by premises personnel. Both of these interventions are ultimately designed to promote responsible beverage service by reducing sales to intoxicated patrons, underage youth, or both. This review assesses the effectiveness of dram shop liability and the enhanced enforcement of overservice laws for preventing excessive alcohol consumption and related harms.

Evidence acquisition: Studies assessing alcohol-related harms in states adopting dram shop laws were evaluated, as were studies assessing alcohol-related harms in regions with enhanced overservice enforcement. Methods previously developed for systematic reviews for the *Guide to Community Preventive Services* were used.

Evidence synthesis: Eleven studies assessed the association of state dram shop liability with various outcomes, including all-cause motor vehicle crash deaths, alcohol-related motor vehicle crash deaths (the most common outcome assessed in the studies reviewed), alcohol consumption, and other alcohol-related harms. There was a median reduction of 6.4% (range of values 3.7% to 11.3% reduction) in alcohol-related motor vehicle fatalities associated with the presence of dram shop liability in jurisdictions where premises are licensed. Other alcohol-related outcomes also showed a reduction. Only two studies assessed the effects of enhanced enforcement initiatives on alcohol-related outcomes; findings were inconsistent, some indicating benefit and others none.

Conclusions: According to *Community Guide* rules of evidence, the number and consistency of findings indicate strong evidence of the effectiveness of dram shop laws in reducing alcohol-related harms. It will be important to assess the possible effects of legal modifications to dram shop proceedings, such as the imposition of statutes of limitation, increased evidentiary requirements, and

From the Community Guide Branch of the Epidemiology and Analysis Program Office (Rammohan, Hahn, Elder, Chattopadhyay, Zometa), National Center for Chronic Disease Prevention and Health Promotion (Brewer), CDC, Atlanta, Georgia; Los Angeles County Department of Health Services (Fielding), Los Angeles, California; Division of Epidemiology and Community Health, University of Minnesota School of Public Health (Toomey), Minneapolis, Minnesota; and the Section of General Internal Medicine, Boston Medical Center (Naimi), Boston, Massachusetts

Author affiliations are from the time the research was conducted.

Names and affiliations of the Task Force members can be found at www.thecommmunityguide.org/about/task-force-members.html.

Address correspondence to: Robert A. Hahn, PhD, MPH, Community Guide Branch, Epidemiology and Analysis Program Office, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop E-69, Atlanta GA 30333. E-mail: rhahn@cdc.gov.

0749-3797/$17.00

doi: 10.1016/j.amepre.2011.06.027

caps on recoverable amounts. According to *Community Guide* rules of evidence, evidence is insufficient to determine the effectiveness of enhanced enforcement of overservice laws for preventing excessive alcohol consumption and related harms.

(Am J Prev Med 2011;41(3):334–343) Published by Elsevier Inc. on behalf of American Journal of Preventive Medicine

Introduction

Excessive alcohol consumption, including binge and underage drinking, is responsible for approximately 79,000 deaths per year in the U.S., making it the third-leading cause of preventable death in the nation.[1] In 1998, the economic cost of excessive drinking was estimated to be $184.6 billion, most of which was due to lost productivity.[2] The reduction of excessive alcohol consumption is thus a matter of major public health and economic interest.

Among a representative sample of U.S. adults from 13 to 14 states, interviewed anonymously by telephone in 2003 and 2004, respectively, 16.3% reported binge drinking (defined as consuming ≥5 drinks per occasion for men, and ≥4 drinks per occasion for women in the past 30 days).[3] Approximately 30% of high school students in the U.S. report binge drinking in the past 30 days,[4,5] and among full-time college students, almost half (48.6%) of male adolescents and 34.4% of female adolescents reported binge drinking.[6]

Among binge-drinking adults, half consumed ≥7 drinks during their most recent drinking episode; 32.7% had their most recent binge episode in a bar/club or restaurant; and between 16.3% and 20.8% of those who drank at a bar or restaurant drove a motor vehicle after binge drinking. Among all episodes of self-reported binge drinking, drinking in a bar/club or restaurant accounted for 54.3% of episodes, compared with 35.7% consumed in a home, and 10% elsewhere.[3] Thus, drinking in bars and restaurants is strongly associated with binge drinking and with alcohol-impaired driving among U.S. adults who report binge drinking. In the U.S., the overservice of alcoholic beverages in on-premises alcohol outlets is a major source of public health problems. Approximately 13,000 U.S. residents die annually from alcohol-related motor vehicle crashes, and many others are injured.[7]

Alcohol control policies have been shown to be effective instruments for preventing alcohol-related harms.[8] Systematic reviews of alcohol policy by the *Guide to Community Preventive Services* (*Community Guide*) have demonstrated the public health benefits of increasing alcohol excise taxes,[9] enhanced enforcement of laws prohibiting alcohol sales to minors,[10] limiting alcohol outlet density,[11] and limiting the days and hours when alcoholic beverages can be sold.[12,13]

This report assesses two law-based interventions for preventing excessive alcohol consumption and related harms, which focus on promoting responsible beverage service in on-premises retail alcohol outlets (e.g., bars or restaurants): dram shop liability and enhanced enforcement of overservice laws. Dram shop liability involves holding the owner or server(s) at an on-premises retail alcohol outlet liable for alcohol-attributable harms (e.g., an alcohol-attributable motor vehicle crash death) caused by a patron who was illegally served alcoholic beverages because the patron was underage or already intoxicated. For there to be liability for service to an intoxicated person, it must be shown that the server either knew or should have known that the patron was intoxicated.

Liability can be established in states either by case law or statute. Most states have enacted dram shop statutes; several states have established dram shop policies by case law/precedent; seven states (Delaware, Kansas, Louisiana, Maryland, Nevada, South Dakota, and Virginia) have neither statutory nor precedent dram shop liability.[14] Most dram shop statutes create barriers to lawsuits not present in common law liability, such as damage caps or stringent evidentiary requirements. Dram shop suits are generally brought by those harmed or by their families. The existence of dram shop liability in a state is thought to promote caution on the part of on-premises owners and staff; owners may purchase liability insurance to protect themselves from financial loss resulting from dram shop law suits.

The second law-based intervention assessed was enhanced enforcement of laws prohibiting "overservice" (defined as sale of alcohol to intoxicated patrons). As with dram shop liability, states vary widely regarding the evidence needed to establish a violation. Enforcement activities are usually carried out by plainclothes or uniformed police, Alcohol Beverage Control personnel, or both. Alcohol Beverage Control Boards are state-operated organizations charged to regulate the sale of alcoholic beverages.

In addition to these direct enforcement actions, this intervention may involve prior notification of retail alcohol outlets of planned enforcement actions, and the training of outlet managers and staff in responsible beverage service, including how to recognize intoxicated patrons and prevent overservice. Legal penalties for overservice may include fines or criminal sanctions for alcohol servers; fines or licensing actions against license holders (including revocation of alcohol sales license); or both. However, a recent systematic review of overservice laws in the U.S. concludes, "The single most notable finding

from the qualitative enforcement research is that enforcement of laws that prohibit alcohol sales to intoxicated patrons is relatively rare."[15]

Both interventions are assumed to work by deterrence, the notion that if premises owners or servers perceive a high probability of incurring substantial penalties by overserving, they will be more likely to avoid doing so. The effect of deterrence depends on three key elements: perceived certainty of detection and punishment, perceived swiftness of punishment, and the perceived severity of punishment.[16] Although both interventions have all the elements, dram shop liability may present a greater perceived threat of meaningful consequences among alcohol premises personnel.[17] This is supported by survey research indicating that actual and perceived threat of dram shop liability are associated with more responsible service practices, but levels of enforcement tend not to be.[18]

Findings, Recommendations, and Directives from Other Reviews and Advisory Groups Related to Dram Shop Liability and Overservice Enforcement

The WHO has published a review that identifies both the enforcement of overservice laws and dram shop liability as effective methods for reducing alcohol-related harms.[8] The Substance Abuse and Mental Health Services Administration's Prevention Enhancement Protocols System[19] recommended "that jurisdictions strictly and uniformly enforce the laws regarding the sale of alcohol to such individuals" (i.e., those who were intoxicated or underage) and "that States and jurisdictions undertake efforts to keep the burden of legal responsibility [for intoxication-related problems caused by patrons] on the owners of drinking establishments and alcohol licensees rather than their employees, such as servers. Jurisdictions might, in fact, consider increasing such liability burdens, not decreasing them."

Evidence Acquisition

The *Community Guide* systematic review process was used to assess whether dram shop liability or overservice law enforcement initiatives lead to decreases in excessive alcohol consumption and related harms. Details of the *Community Guide* review process are presented elsewhere.[20,21] The process involves forming a systematic review team; developing a conceptual approach to organizing, grouping, and selecting interventions; prioritizing these interventions; systematically searching for and retrieving the existing research evidence on the effects of the interventions; abstracting information from each study that meets qualifying criteria; assessing the quality of each study; drawing conclusions about the body of evidence on intervention effectiveness; and translating the evidence on effectiveness into recommendations.

The systematic review team consists of systematic review methodologists and subject matter experts from a range of agencies, organizations, and academic institutions. The review team works under the oversight of the nonfederal, independent Task Force on Community Preventive Services (Task Force), which directs the work of the *Community Guide*.

The systematic review team (the team) collects and summarizes evidence on (1) effectiveness of interventions in improving health-related outcomes of interest and (2) additional benefits and potential harms of the intervention on other health and nonhealth outcomes. When an intervention is shown to be effective, information is also included about (3) the applicability (i.e., generalizability) of the evidence to diverse population segments and settings, (4) the economic impact of the intervention, and (5) barriers to implementation. Such information may also be provided in the absence of sufficient evidence of effectiveness. The team then presents the results of this review process to the Task Force, which determines whether all of the evidence presented is sufficient to warrant a recommendation for practice or policy.[20]

The rules of evidence under which the Task Force makes its determination address several aspects of the body of evidence, including the number of studies of different levels of design suitability and execution, the consistency of the findings among studies, the public health importance of the overall effect size, and the balance of benefits and harms of the intervention.

Conceptual Approach and Analytic Framework

Dram shop liability. The effect of dram shop liability on alcohol-related outcomes may be influenced by at least two factors (Figure 1). In states with dram shop liability, premises owners perceive a potential for liability suits,[22] which may be communicated by public and trade media and word of mouth. Such perception may result in increased training of outlet personnel in responsible beverage service, increased motivation, increased oversight, and increased compliance with liquor laws.

These changes may result in reduced illegal beverage service, including service to intoxicated patrons and underage youth, and ultimately reduced excessive consumption and related harms. In states with dram shop liability, many owners of on-premises outlets commonly purchase liability insurance and thus have some financial protection against possible legal action. Because of this protection, insurance may reduce the deterrent effect of liability; the team encountered no evidence regarding this conjecture.

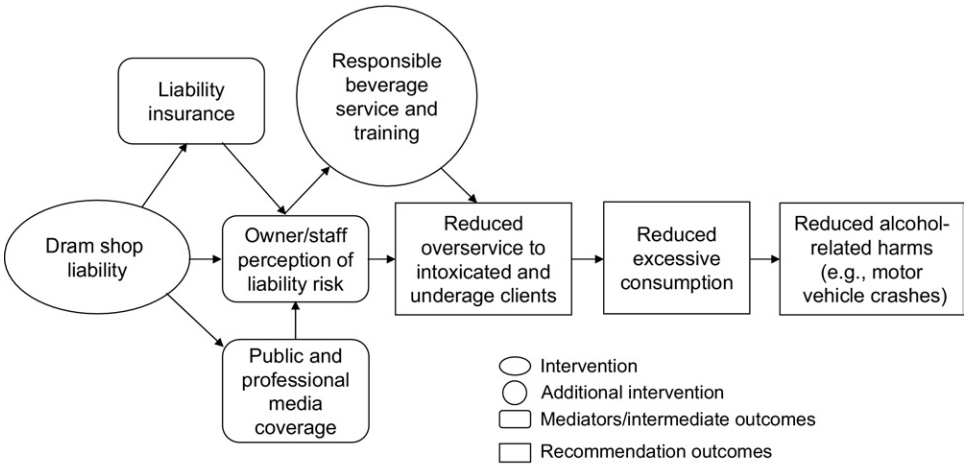


Figure 1. Analytic framework: dram shop liability

Enhanced enforcement of overservice laws. Overservice enforcement initiatives are designed to increase the perceived risk by servers and managers of on-premises retail alcohol outlets of sanctions resulting from serving intoxicated patrons (Figure 2). In response to such initiatives, establishment personnel may undergo training to improve the ability of servers to detect patrons who are intoxicated, so that they can then refuse to serve additional alcohol. These intermediate consequences of enhanced enforcement are hypothesized to reduce excessive alcohol consumption and alcohol-related harms.

Inclusion Criteria

To qualify as a candidate for inclusion in this review, a study had to:

- Evaluate the effectiveness of dram shop liability or initiatives for enhanced enforcement of overservice regulations that could and did apply legal or administrative sanctions.
- Be conducted in a country with a high-income economy,[a,23] be primary research (rather than a review of other research), and be published in English.
- Compare attributes of participants before and after the implementation of the intervention or compare a group receiving the intervention with a group not receiving it.
- Report outcomes related to excessive alcohol consumption or related harms, such as alcohol-impaired driving or alcohol-related motor vehicle crashes.

Search for Evidence

The following databases were searched for this review: CINAHL, EconLit, Embase, ERIC (CSA), NLM Gateway, NTIS (National Technical Information Service), PsycINFO, PsycNET (APA), MEDLINE, Science Direct, Social Services Abstracts, and Sociological Abstracts (CSA) Web-of-Science. All publication years covered in each database were searched up to October 2007. (Details of the search strategy are available at www.thecommunityguide.org/alcohol/supportingmaterials/SSalcoholuse.html.) Reference lists of articles reviewed as well as lists in review articles were also searched, and subject matter experts consulted for additional references. Published government reports were included, but not unpublished manuscripts because they are not commonly available in the public domain.

Evidence Synthesis

Assessing the Quality and Summarizing the Body of Evidence on Effectiveness

Each study that met the inclusion criteria for candidate studies was read by two reviewers who used standardized criteria (www.thecommunityguide.org/about/methods.html) to assess the suitability of the study design and threats to validity. Uncertainties and disagreements between the reviewers were reconciled by consensus among team members. Classification of the designs of reviewed studies accords with the way in which study findings were used in the review and with the standards of the *Community Guide* review process[21]; they may differ from the classification reported in the original studies.

Each candidate study for this review was evaluated for quality of study design and execution. Studies with greatest design suitability were those in which outcome data on exposed and comparison populations were collected prospectively, such as panel (i.e., cohort) studies; studies with moderate design suitability were those in which data on exposed and comparison populations were collected retrospectively or in which there were multiple pre- or post-intervention measurements, but no concurrent

[a]Countries with high-income economies as defined by the World Bank are Andorra, Antigua and Barbuda, Aruba, Australia, Austria, The Bahamas, Bahrain, Barbados, Belgium, Bermuda, Brunei Darussalam, Canada, Cayman Islands, Channel Islands, Cyprus, Denmark, Estonia, Faeroe Islands, Finland, France, French Polynesia, Germany, Greece, Greenland, Guam, Hong Kong (China), Iceland, Ireland, Isle of Man, Israel, Italy, Japan, Republic of Korea, Kuwait, Liechtenstein, Luxembourg, Macao (China), Malta, Monaco, Netherlands, Netherlands Antilles, New Caledonia, New Zealand, Norway, Portugal, Puerto Rico, Qatar, San Marino, Saudi Arabia, Singapore, Slovenia, Spain, Sweden, Switzerland, Trinidad and Tobago, United Arab Emirates, United Kingdom, U.S., Virgin Islands (U.S.).

comparison population; and studies with least suitable designs were cross-sectional studies or those that had no separate comparison population and only a single pre- and post-measurement in the intervention population.

On the basis of the number of threats to validity—such as poor measurement of exposure or outcome, lack of control of potential confounders, or high attrition—studies were characterized as having good (≤1 threat to validity); fair (2–4 threats); or limited (≥5 threats) quality of execution. Studies with good or fair quality of execution and any level of design suitability (greatest, moderate, or least) qualified for the body of evidence; studies with limited quality of execution were excluded.

With dram shop liability jurisdiction as the unit of analysis, effect estimates were calculated as relative percentage change using the following formulas:

- For studies with before-and-after measurements and concurrent comparison groups:

$$\text{Effects Estimate} = [(I_{post}/C_{post})/(I_{pre}/C_{pre}) - 1] \times 100\,\%,$$

where:

I_{post} = last reported outcome rate in the intervention group after the intervention;

I_{pre} = last reported outcome rate in the intervention group before the intervention;

C_{post} = last reported outcome rate in the comparison group after the intervention;

C_{pre} = last reported outcome rate in the comparison group before the intervention.

- For studies with before-and-after measurements but no concurrent comparison:

$$\text{Effects Estimate} = [(I_{post}/I_{pre})/I_{pre}] \times 100\%.$$

When appropriate data were provided, CIs for effect estimates were calculated. When a body of evidence included three or more studies, medians and data range of values were reported.

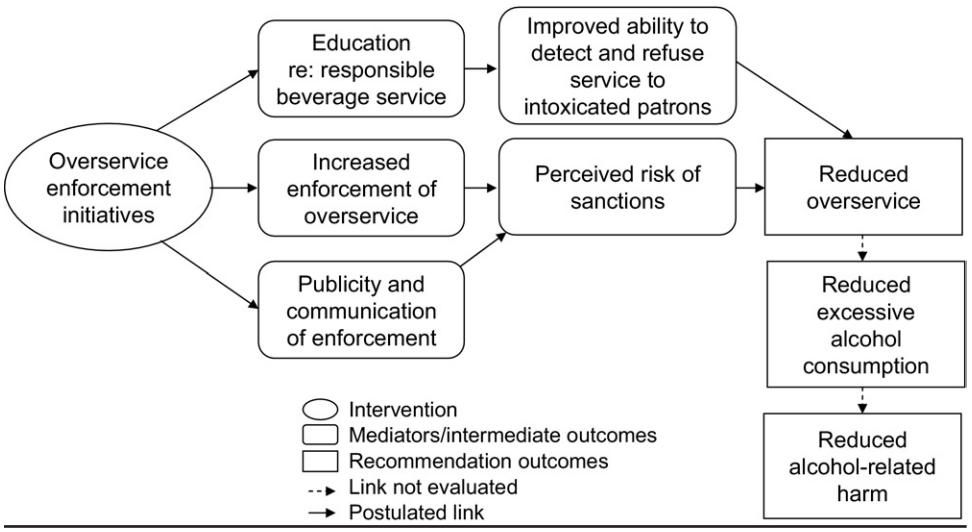


Figure 2. Analytic framework: enhanced enforcement of alcohol overservice initiatives

Intervention Effectiveness: Dram Shop Liability

The review included 11 studies[22,24-33] of the effectiveness of dram shop liability in preventing excessive alcohol consump-

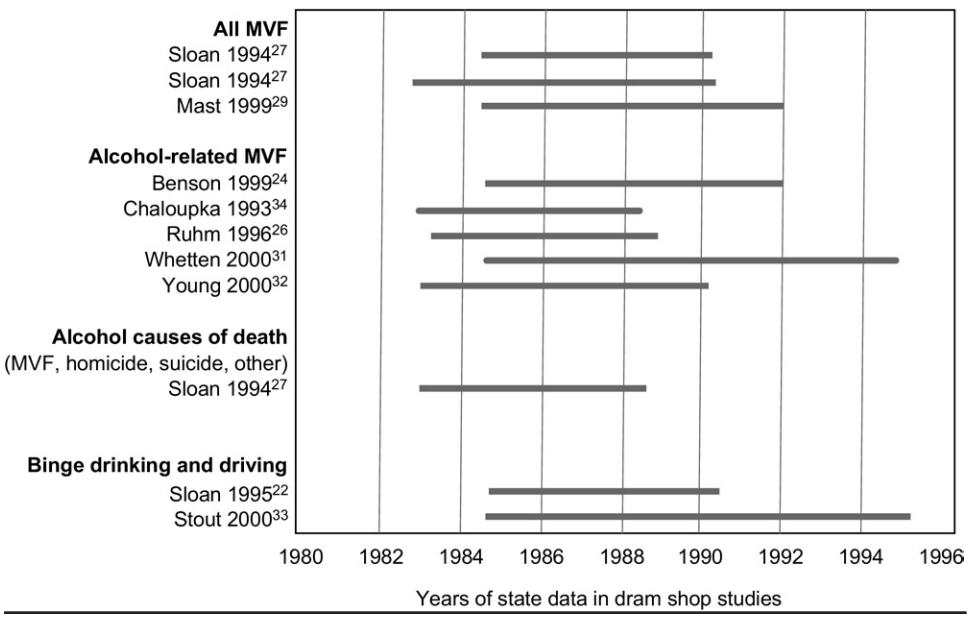


Figure 3. Overlap in time periods among dram shop studies indicating dependence among the data
MVF, motor vehicle fatalities

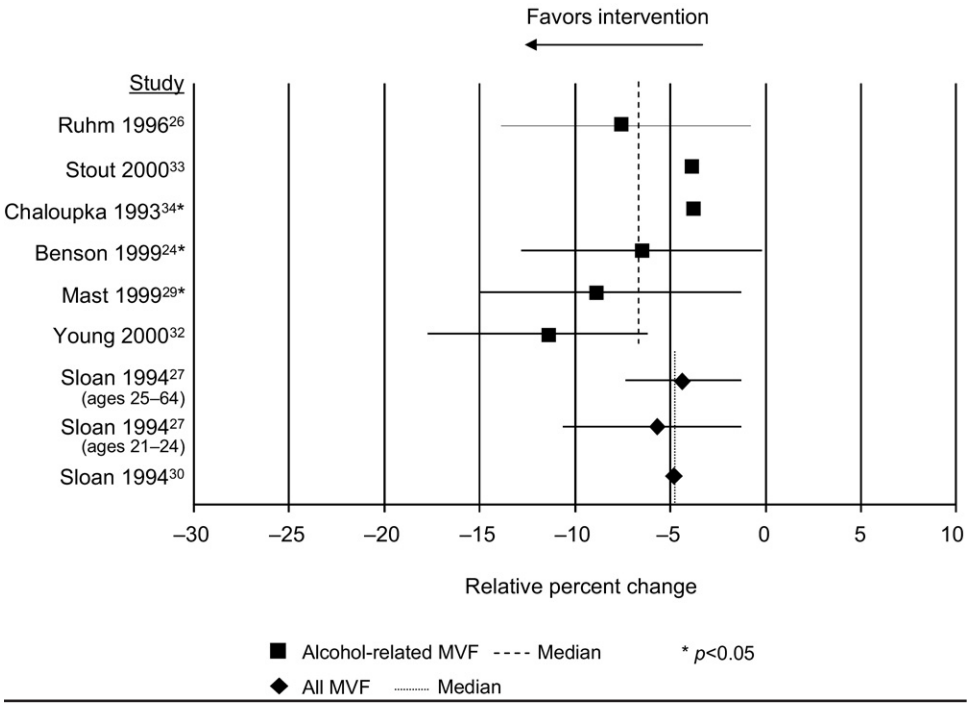


Figure 4. Relative percentage change in alcohol-related and all-cause motor vehicle fatalities associated with dram shop liability

tion and related harms. Nine[24-32] of these were of greatest design suitability and two[22,33] were of least suitable design. Five[24-28] were of good quality of execution and six[22,29-33] were of fair quality of execution.

All studies but one[28] were panel studies of U.S. states using econometric models to assess the effects of dram shop liability and other interventions on diverse outcomes. These studies covered overlapping time periods (Figure 3) and states, and thus are not entirely independent. However, the models in these studies included different covariates and assessed effects on different outcomes.

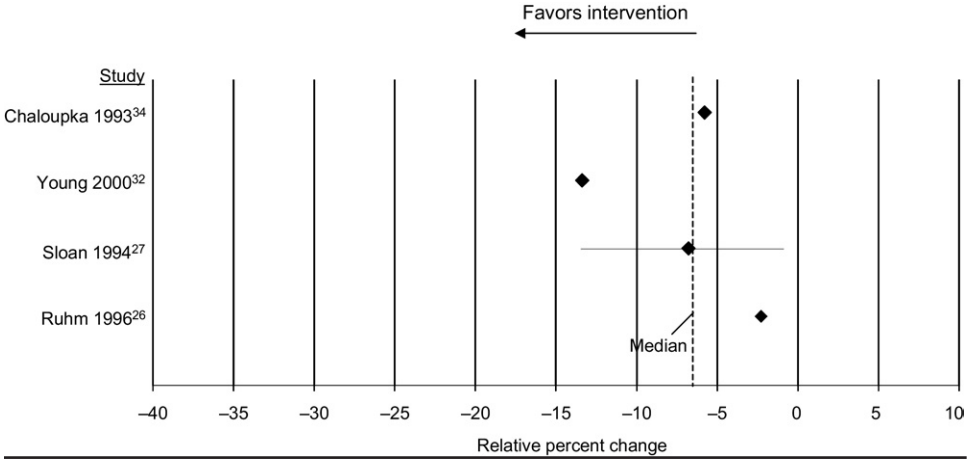


Figure 5. Relative percentage change in alcohol-involved motor vehicle fatalities associated with dram shop liability for underage drinkers

Six studies[24-26,29,32,33] that assessed the effects of dram shop liability on alcohol-related motor vehicle fatalities found a median reduction of 6.4% (range of values 3.7% to 11.3% reduction) associated with these policies (Figure 4). Two studies[27,30] that assessed all-cause motor vehicle fatalities (in which not all crashes were attributable to alcohol) found a median reduction of 4.8% (Figure 4). Those[26,27,32,34] that reported all-cause motor vehicle fatalities among underage drinkers all found reductions of between 2.2% and 13.0% (Figure 5). Only two studies[22,33] assessed changes in alcohol consumption (i.e., self-reported binge drinking) as an outcome; both found small, nonsignificant decreases (1.2% and 2.4%) associated with dram shop liability in states (data not shown in graphic).

One panel study[28] assessed the effects of two lawsuits brought against on-premises alcohol outlets in Texas in 1983 and 1984. These suits, filed by the families of people killed in alcohol-related motor vehicle crashes, were publicized in general public and alcohol-industry periodicals. The researchers used interrupted time-series models to assess the effects of these suits on single-vehicle nighttime crashes in Texas, compared with the 47 other contiguous U.S. states. The researchers found that the first suit was associated with a 6.6% decrease (95% CI=0.5%, 11.3%) in single-vehicle nighttime crashes, and the second suit was associated with an additional 5.3% decrease (95% CI=0.4%, 10.1%) in single-vehicle nighttime crashes (data not shown).

One study[27] reported reductions in rates of suicide, homicide, and alcohol-

related medical conditions; the last two findings were significant ($p<0.01$), the first was not ($p>0.10$).

Summary: Results of Dram Shop Liability Review

Eleven studies of dram shop liability consistently found that this intervention reduced motor vehicle crash deaths in general and alcohol-related crash deaths in particular. Strong evidence indicated that dram shop liability is an effective intervention for reducing alcohol-related harms, as indicated by reduced motor vehicle crashes. Two studies of binge drinking did not provide sufficient independent evidence on the effect of dram shop liability on excessive consumption.

Intervention Effectiveness: Enhanced Enforcement of Overservice Laws

Two studies assessed the effects of enhanced enforcement of overservice laws.[35,36] Both studies were of greatest design suitability and fair quality of execution. Both studies had pre–post designs, with concurrent comparisons. They provided information on two key components of enhanced enforcement initiatives: (1) owners and servers of establishments serving alcohol were made aware of enforcement efforts through media coverage, letters, presentations, and reports containing information about undercover visits; and (2) owners and servers were offered server training and educational materials on responsible beverage service.

One study[35] assessed an overservice initiative in Washtenaw County, Michigan. State Police Department enforcement was countywide, and was followed by after-visit reports notifying licensees visited by undercover police officers that enforcement was in progress. After 1 year, prevalence of service to pseudo-intoxicated pseudo-patrons in intervention settings decreased by 14.8% compared with control settings, and the percentage of driving under the influence (DUI) arrestees in the experimental county who reported having consumed their last drink in an intervention setting decreased by a relative 36.3% in the intervention sites.

The second study[36] assessed the impact of enhanced enforcement of overservice laws administered by the Washington State Liquor Control Board in bars and restaurants in Washington State, assessing service to pseudo-patrons and DUI associated with alcohol consumption in intervention and control outlets. Compared with control sites, alcohol sales to pseudo-intoxicated pseudo-patrons in intervention sites increased 82.6%, and the average number of monthly DUI arrests in which intervention establishments were identified as "place of last drink" decreased by 31.2% ($p>0.05$).

Summary of Results of Enhanced Enforcement Initiative Review

The only available two studies of enhanced enforcement of overservice laws included in this review differed substantially in design (i.e., sample size and analysis) and findings. All outcomes in the Michigan study[35] had favorable and significant findings, but the Washington study[36] had contrary results (i.e., an apparent increase in service to pseudo-intoxicated pseudo-patrons, but an apparent decrease in DUIs). The small number of studies and inconsistent findings provided an insufficient body of evidence to determine the effectiveness of enhanced enforcement of overservice laws on excessive alcohol consumption and related harms.

Potential Harms, Additional Benefits, and Barriers to Implementation

Although dram shop liability appears to have deterrent effects, litigation may be an expensive and inefficient method of achieving this outcome. Under dram shop liability, for legal actions to be brought against a manager or server in an on-premises establishment, there must have been both illegal beverage service (e.g., service to an intoxicated patron) and harm to someone as a result of this illegal service. In addition, an individual who experienced harms related to illegal beverage service (or a representative of that individual) must prove that illegal service took place, which may be difficult.

On the other hand, dram shop liability can foster an environment that encourages responsible server behavior, and thus encourages investment in server training and other primary prevention strategies. This intervention can also help to create a retail environment that makes responsible beverage service the norm and, thus, does not unfairly disadvantage responsible beverage servers.

Despite these challenges to implementation, dram shop liability may be useful because it focuses on a regulated environment which is thus amenable to control. Furthermore, on-premises alcohol outlets have been strongly associated with high-intensity binge drinking (i.e., a higher self-reported number of drinks per binge episode) and related risk behaviors, such as driving after binge drinking, furnishing a strong public health justification for targeting interventions to these settings.

One harm that may be posited with overservice enforcement is that underage drinkers and intoxicated patrons in on-premises facilities may move to uncontrolled settings to consume additional alcohol. The team found no evidence on this issue.

Applicability

Much of the research assessing the effectiveness of dram shop liability was conducted before the enactment of various

caps on the financial liability of servers and managers in dram shop cases in the late 1990s. These changes may have modified the effectiveness of this intervention, limiting its applicability to current circumstances. Some states have instituted statutes of limitation that require injured plaintiffs to sue within a specified time period. The standards of evidence required in dram shop liability cases have also grown more stringent, making it increasingly difficult to prove illegal beverage service. In addition, knowledge of and access to legal services vary greatly by SES, making it difficult for some segments of the population to obtain legal services for dram shop litigation.

Economic Efficiency

Dram Shop Liability

The systematic economic review did not identify any studies that examined the costs and benefits of dram shop liability. Thus an economic analysis was not possible for this review.

Overservice Law Enforcement

Although insufficient evidence to determine the effectiveness of enhanced enforcement of overservice law initiatives was found, the systematic economic review identified two analyses that estimated the costs of enhanced enforcement of overservice laws, and found substantial benefit. Both studies were based on the findings of the Michigan program described above.[35] This evidence is summarized now, in case evidence accrues in the future to support this intervention.

Levy and Miller[37] conducted a cost–benefit analysis and estimated that the combined police, supervisory, and miscellaneous costs for enhanced enforcement of overservice laws in Ann Arbor city and Washtenaw County was $84,296 in 2009 dollars. The estimated benefits of the program (in 2009 dollars) attributable to reduced tavern-related DUI cases alone were approximately $800,000 in medical cost savings, and about $6.1 million when additional savings were taken into account (e.g., reduced demand for emergency services, such as fire and police; travel delays for motorists who would have otherwise been involved in crashes; property damage; costs to employers caused by workplace disruption; productivity losses for employees; and administrative costs, including claims-processing and legal and court costs).

If pain and suffering and lost quality of life were added, the economic benefit of enhanced enforcement would increase to about $16.6 million. However, the estimated benefit of this intervention would decline to about $8 million if only averted external costs (i.e., costs to third parties) were included. In the best-case scenario, when all societal benefits from averted DUI crashes are properly accounted for, each dollar invested in the program returned more than $196 in benefits. It is important to note that this benefit–cost estimate is exclusively based on traffic-related injuries, and does not consider other household injuries (e.g., assault and domestic violence) resulting from excessive drinking. The benefits of overservice enforcement may be reduced if people who are refused a drink in one location resume drinking at another location and then drive or engage in other risky behavior.

McKnight et al.[35] extended the cost–benefit analysis of the Washtenaw County Service to Intoxicated Patrons program to the national level, using benefit estimates from Miller and Levy.[37] Assuming an estimated total cost of a nationwide law enforcement effort of $74.5 million per year and annual net savings of $21 billion from averted costs related to DUIs and crashes, they reported a benefit of $282 for each dollar invested in the program. (All dollar figures are adjusted to 2009 dollars based on the consumer price index.) Thus, studies indicate the large potential cost benefit of this intervention, were it found to be effective.

Research Gaps

As noted, many of the studies included in this review were conducted prior to the enactment in the late 1990s of various caps on financial liability of servers and managers in dram shop cases, in addition to statutes of limitation and increased legal evidence requirements. Further research is needed to assess what impact, if any, these limits on liability have had on the effectiveness of dram shop laws in reducing excessive alcohol consumption and related harms.

Additional studies are needed to assess how effective enhanced enforcement of overservice regulations is in reducing excessive alcohol consumption and related harms. It would be useful to ascertain barriers to effective enforcement. In addition, research is needed to assess the role of the media in publicizing enhanced enforcement and enhancing its effectiveness and the potential role of responsible beverage service training programs in reducing overservice and thus enhancing the effectiveness of enforcement. The latter is important as these multicomponent server intervention programs may prove beneficial in decreasing excessive alcohol consumption and related harms in on-premises retail alcohol settings. The potential cost savings to owners of on-premises retail alcohol outlets through the promotion of responsible beverage service will be useful in assessing economic benefits.

The signs of intoxication that a patron exhibits may be difficult for servers or law enforcement officials to identify. To help servers avoid engaging in illegal service, additional research is needed to improve methods for identifying patrons who are intoxicated, underage, or

both. Other methods for avoiding overservice can also be explored, such as counting drinks and spacing out the frequency of drink service through the use of food or nonalcoholic drinks after a predetermined threshold has been achieved.

Although enforcement of existing laws and regulations prohibiting service of alcohol to intoxicated patrons appears to be cost-beneficial based on the estimated findings from Washtenaw County, additional studies are needed for a more reliable estimate of the economic value of enforcement.

Finally, additional research is needed to assess the effectiveness of both dram shop liability and enhanced enforcement in achieving these broader societal impacts, but if further research corroborates these findings, dram shop liability and enhanced enforcement of overservice laws could provide many collateral benefits.

Discussion

This review assesses two law-based approaches to promoting responsible beverage service in on-premises retail alcohol settings, including bars and restaurants. Room and colleagues[38] have argued that "the general rule in such situations is that it is easier and more effective for the state to influence licensed occupational behavior than it is to influence the behavior of private customers." Evidence of the effectiveness of one of these approaches—dram shop liability—is strong. The effectiveness of this approach, however, may be diminished by restrictions on these laws by stringent monetary caps, evidentiary requirements, and statutes of limitations. The effects of these restrictions should be investigated.

No studies of the economic effects of dram shop liability were found. Nevertheless, given the association between drinking in on-premises retail alcohol outlets and high-intensity binge drinking, and the relationship between binge drinking and a variety of other health and social problems, including alcohol-impaired driving and interpersonal violence, the potential economic impact of promoting more responsible beverage service by holding managers and servers responsible for harms resulting from illegal beverage service could be substantial. The real benefit of maintaining strong dram shop liability, however, may result from creating a business environment that supports responsible beverage service at on-premises retail outlets without penalizing servers and mangers who strive to comply with liquor control laws. This would, in turn, help to reduce illegal beverage service and harms resulting from it, thus decreasing the likelihood that dram shop liability cases will need to be adjudicated by the courts. Furthermore, reduced prevalence of drunken and disorderly conduct in on-premises retail alcohol outlets may also reduce the cost of operations for owners, and thus offset the potential loss in sales that may result from the intervention.[37,39]

Effectiveness of the other approach assessed in this review—the enhanced enforcement of overservice regulations—could not be determined because there were too few studies and inconsistent findings. There are examples of intensive enforcement efforts among U.S. states. New Mexico has one of the strongest overservice enforcement programs. As of 2006, alcohol licensees in New Mexico are subject to license suspension for a first violation and to license revocation following three violations within 1 year. Presumptive evidence of overservice can be established by a blood alcohol level of 0.14 g/dL in patrons within 90 minutes of exiting a drinking establishment. In addition, the New Mexico Department of Public Safety has a "Mobile Strike Team" that investigates licensed establishments where overservice has been noted.

Awareness of such policies may increase the level of deterrence in the state. A National Highway Traffic Safety Administration report, "Laws Prohibiting Alcohol Sales to Intoxicated Persons,"[15] proposes a series of "best practices," including the use of presumptive evidence of a blood alcohol level of 0.14 g/dL in patrons exiting a drinking establishment, as in New Mexico; enactment of service to intoxicated patron legislation in all states; the collection of data for the monitoring of alcohol-related harms; and the training of law enforcement personnel in the enforcement of service to intoxicated patron rules. Such measures may assist in the development of more-effective procedures for the reduction of the harms associated with excessive alcohol consumption in the U.S.

The authors gratefully acknowledge the contributions to this review by Ralph Hingson, ScD, MPH (NIAAA), Steve Wing (SAMHSA), and Jim Mosher, JD (Alcohol Policy Consultations).

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC.

No financial disclosures were reported by the authors of this paper.

References

1. CDC. Alcohol-attributable deaths and years of potential life lost—U.S., 2001. MMWR Morb Mort Wkly Rep 2004;53(37):866–70.
2. Harwood H. Updating estimates of the economic costs of alcohol abuse in the U.S.: estimates, update methods, and data. Report prepared by the Lewin Group for the National Institute on Alcohol Abuse and Alcoholism 2000.

3. Naimi TS, Nelson DE, Brewer RD. Driving after binge drinking. Am J Prev Med 2009;37(4):314–20.
4. National Center for Health Statistics. Health, U.S., 2005 with chartbook on trends in the health of America, Issue 1232. 2005.
5. Miller JW, Naimi TS, Brewer RD, Jones SE. Binge drinking and associated health risk behaviors among high school students. Pediatrics 2007;119(1):76–85.
6. Johnston LD, O'Malley PM, Bachman JG, Schulenberg JE. Monitoring the Future: national survey results on drug use, 1975–2008. Volume II, College students and adults ages 19–50. Bethesda MD: National Institute on Drug Abuse, 11-1-2009. NIH Publication No. 09-7403.
7. Naumann RB, Dellinger AM, Zaloshnja E, Lawrence BA, Miller TR. Incidence and total lifetime costs of motor vehicle–related fatal and nonfatal injury by road user type, U.S., 2005. Traffic Inj Prev 2010; 11(4):353–60.
8. Babor T, Caetano R, Casswell S, et al. Alcohol: no ordinary commodity—research and public policy. 2nd edition. Oxford, UK: Oxford University Press, 2010.
9. Task Force on Community Preventive Services. Increasing alcohol beverage taxes is recommended to reduce excessive alcohol consumption and related harms. Am J Prev Med 2010;38(2):230–2. www.thecommunityguide.org/alcohol/IncreasingAlcoholBeverageTaxesRecommendedReduceExcessiveAlcoholConsumptionRelatedHarms.pdf.
10. Guide to Community Preventive Services. Reducing excessive alcohol use: enhanced enforcement of laws prohibiting sales to minors. www.thecommunityguide.org/alcohol/lawsprohibitingsales.html.
11. Task Force on Community Preventive Services. Recommendations for reducing excessive alcohol consumption and alcohol-related harms by limiting alcohol outlet density. Am J Prev Med 2009;37(6):570–1. www.thecommunityguide.org/alcohol/RecommendationsReducingExcessiveAlcoholConsumptionAlcohol-RelatedHarmsLimitingAlcoholOutletDensity.pdf.
12. Hahn R, Kuzara J, Elder R, et al.; Task Force on Community Preventive Services. The effectiveness of policies restricting hours of alcohol sales in preventing excessive alcohol consumption and related harms. Am J Prev Med 2010;39(6):590–604. www.thecommunityguide.org/alcohol/EffectivenessofPoliciesRestrictingHoursofAlcoholSalesinPreventingExcessiveAlcoholConsumptionandRelatedHarms.pdf.
13. Middleton J, Hahn R, Kuzara J, et al.; Task Force on Community Preventive Services. Effectiveness of policies maintaining or restricting days of alcohol sales on excessive alcohol consumption and related harms. Am J Prev Med 2010;39(6):575–89. www.thecommunityguide.org/alcohol/EffectivenessofPoliciesMaintainingRestrictionDaysofAlcoholSalesonExcessiveAlcoholConsumptionandRelatedHarms.pdf.
14. National Highway Traffic Safety Administration. Digest of impaired driving and selected beverage control laws. Twenty-fourth edition. www.nhtsa.gov/DOT/NHTSA/Traffic Injury Control/Articles/AssociatedFiles/HS810827.pdf.
15. Mosher J, Hauck A, Carmona M, et al. Legal research report: laws prohibiting alcohol sales to intoxicated persons. www.nhtsa.gov/DOT/NHTSA/Traffic%20Injury%20Control/Articles/Associated%20Files/811142.pdf.
16. Kleiman M, Kilmer B. The dynamics of deterrence. Proc Natl Acad Sci U S A 2009;106(34):14230–5.
17. Gehan JP, Toomey TL, Jones-Webb R, Rothstein C, Wagenaar AC. Alcohol outlet workers and managers: focus groups on responsible service practices. J Alcohol Drug Educ 1999;44(2):60–71.
18. Liang L, Sloan FA, Stout EM. Precaution, compensation, and threats of sanction: the case of alcohol servers. Int Rev Law Econ 2004; 24(1):49–70.
19. Grover PL, Bozzo B, Giesbrecht N, et al. Preventing problems related to alcohol availability: Environmental approaches. 2010. DHHS Pub. No. SMA 99-3298.
20. Briss PA, Zaza S, Pappaioanou M, et al. Developing an evidence-based Guide to Community Preventive Service—methods. Am J Prev Med 2000;18(1S):35–43. www.thecommunityguide.org/methods/methods-ajpm-developing-guide.pdf.
21. Zaza S, Wright-De Aguero LK, Briss PA, et al. Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Am J Prev Med 2000;18(1S):44–74.www.thecommunityguide.org/library/ajpm355_d.pdf.
22. Sloan FA, Reilly BA, Schenzler C. Effects of tort liability and insurance on heavy drinking and drinking and driving. J Law Econ 1995; 38(1):49–77.
23. World Bank. World development indicators 2006. devdata.worldbank.org/wdi2006/contents/cover.htm.
24. Benson BL, Rasmussen DW, Mast BD. Deterring drunk driving fatalities: an economics of crime perspective. Int Rev Law Econ 1999; 19:205–25.
25. Chaloupka FJ, Saffer H, Grossman M. Alcohol-control policies and motor-vehicle fatalities. J Legal Stud 1993;22(1):161–86.
26. Ruhm CJ. Alcohol policies and highway vehicle fatalities. J Health Econ 1996;15(4):435–54.
27. Sloan FA, Reilly BA, Schenzler CM. Tort liability versus other approaches for deterring careless driving. Int Rev Law Econ 1994;14(1): 53–71.
28. Wagenaar AC, Holder HD. Effects of alcoholic beverage server liability on traffic crash injuries. Alcohol Clin Exp Res 2007;15(6):942–7.
29. Mast BD, Benson BL, Rasmussen DW. Beer taxation and alcohol-related traffic fatalities. S Econ J 1999;66(2):214–7.
30. Sloan FA, Reilly BA, Schenzler C. Effects of prices, civil and criminal sanctions, and law enforcement on alcohol-related mortality. J Stud Alcohol 1994;55(4):454.
31. Whetten-Goldstein K, Sloan FA, Stout E, Liang L. Civil liability, criminal law, and other policies and alcohol-related motor vehicle fatalities in the U.S.: 1984–1995. Accid Anal Prev 2000;32(6):723–33.
32. Young DJ, Likens TW. Alcohol regulation and auto fatalities. Int Rev Law Econ 2000;20(1):107–26.
33. Stout EM, Sloan FA, Liang L, Davies HH. Reducing harmful alcohol-related behaviors: effective regulatory methods. J Stud Alcohol 2000;61(3):402–12.
34. Chaloupka FJ, Saffer H, Grossman M. Alcohol-control policies and motor-vehicle fatalities. J Legal Stud 1993;22:161.
35. McKnight AJ, Streff FM. The effect of enforcement upon service of alcohol to intoxicated patrons of bars and restaurants. Accid Anal Prev 1994;26(1):79–88.
36. Ramirez R, Nguyen D, Cannon C, Carmona M, Freisthler BA. A campaign to reduce impaired driving through retail-oriented enforcement in Washington State. www.nhtsa.gov/DOT/NHTSA/Traffic%20Injury%20Control/Articles/Associated%20Files/810913.pdf.
37. Levy DT, Miller TR. A cost-benefit analysis of enforcement efforts to reduce serving intoxicated patrons. J Stud Alcohol 1995;56:240–7.
38. Room R, Babor T, Rehm J. Alcohol and public health. Lancet 2005;365:519–30.
39. Sloan F, Chepke L. The law and economics of public health. Foundations and trends in microeconomics. Law Econ Pub Health 2007;3(5–6):391–430.

SUBSTANCE ABUSE TREATMENT, PREVENTION, AND POLICY

RESEARCH Open Access

Chronic disease and recent addiction treatment utilization among alcohol and drug dependent adults

Sharon Reif[1*], Mary Jo Larson[1], Debbie M Cheng[2], Donald Allensworth-Davies[2], Jeffrey Samet[2,3,4] and Richard Saitz[2,3,4]

Abstract

Background: Chronic medical diseases require regular and longitudinal care and self-management for effective treatment. When chronic diseases include substance use disorders, care and treatment of both the medical and addiction disorders may affect access to care and the ability to focus on both conditions. The objective of this paper is to evaluate the association between the presence of chronic medical disease and recent addiction treatment utilization among adults with substance dependence.

Methods: Cross-sectional secondary data analysis of self-reported baseline data from alcohol and/or drug-dependent adults enrolled in a randomized clinical trial of a disease management program for substance dependence in primary care. The main independent variable was chronic medical disease status, categorized using the Katz Comorbidity Score as none, single condition of lower severity, or higher severity (multiple conditions or single higher severity condition), based on comorbidity scores determined from self-report. Asthma was also examined in secondary analyses. The primary outcome was any self-reported addiction treatment utilization (excluding detoxification) in the 3 months prior to study entry, including receipt of any addiction-focused counseling or addiction medication from any healthcare provider. Logistic regression models were adjusted for sociodemographics, type of substance dependence, recruitment site, current smoking, and recent anxiety severity.

Results: Of 563 subjects, 184 (33%) reported any chronic disease (20% low severity; 13% higher severity) and 111 (20%) reported asthma; 157 (28%) reported any addiction treatment utilization in the past 3 months. In multivariate regression analyses, no significant effect was detected for chronic disease on addiction treatment utilization (adjusted odds ratio [AOR] 0.88 lower severity vs. none, 95% confidence interval (CI): 0.60, 1.28; AOR 1.29 higher severity vs. none, 95% CI: 0.89, 1.88) nor for asthma.

Conclusions: In this cohort of alcohol and drug dependent persons, there was no significant effect of chronic medical disease on recent addiction treatment utilization. Chronic disease may not hinder or facilitate connection to addiction treatment.

Keywords: addiction, substance abuse, substance abuse, treatment, medical care, chronic disease

Introduction

Chronic medical diseases are long-lasting conditions, often progressive, and often controllable with continuing care and behavior change. In an era of increasing health care costs [1], chronic disease stands out as a major contributor [2] with alcohol and drug use disorders playing an exacerbating role [3-5]. To improve population health and reduce U.S. health care costs, particular attention must be paid to the role of chronic diseases, including medical, psychiatric, and substance use disorders.

Substance use disorders (SUDs) are prevalent in about 9% of the U.S. population, but only 10% of those individuals access addiction treatment [6]. Although medical needs were not explicitly mentioned in this household

* Correspondence: reif@brandeis.edu
[1]Institute for Behavioral Health, Heller School for Social Policy and Management, Brandeis University, 415 South Street, MS035, Waltham, MA 02454, USA
Full list of author information is available at the end of the article

survey as a reason for not accessing addiction treatment [6], medical needs may interfere with other priorities. For example, chronic medical conditions could cause functional limitations that preclude access to other health care. Interference from pressing medical conditions may thus contribute to perceived unimportance of addiction treatment among the 94% of people [6] deemed in need but who reported they did not need it. To enhance care of both chronic medical diseases and addiction, it is important to better understand their relationships.

More than 40% of the U.S. population is estimated to have at least one chronic medical disease [2,7], with prevalence higher for women and increasing dramatically with age [7]. Chronic medical disease is related to 70% of all U.S. deaths [2], and accounts for more than 75% of health care costs [2] and high out of pocket health care spending [7]. Diabetes, hypertension, and high cholesterol comprise nearly a third of all chronic conditions [7], and just seven conditions account for about one-fourth of annual outpatient visits and hospital discharges [8].

Chronic disease may have unexpected consequences on healthcare utilization, potentially reducing treatment of unrelated disorders, perhaps due to less time available, complexity of navigating multiple problems at once, or a preference to focus on the most troublesome problem [9]. In practice, however, perhaps due to more frequent interaction with medical providers, mental disorders are more often detected among patients with chronic medical diseases than among other patients [10]. Physician awareness of a substance abuse problem also is more likely when the person has episodic or chronic medical illness, however SUDs remain undiagnosed by primary care physicians in nearly half of patients seeking addiction treatment [11].

It is well documented that SUDs cause or exacerbate certain medical conditions and increase their costs [3,12-16]. Incidence for many chronic medical diseases, including hypertension, diabetes, asthma, chronic liver disease, chronic obstructive pulmonary disease (COPD), pain, and stroke is elevated, and for some disorders more than doubled, among people with SUDs [17-20]. Many more conditions are worsened by, have their management impaired by, or are attributable to SUDs [21]. Chronic conditions are prevalent even in younger individuals with SUDs [22]. These findings highlight the importance of medical management as an essential element of addiction treatment [23] particularly in early recovery [24].

Several examples of this complex interplay are reported in the literature. Alcohol use disorders increased medical illness complexity in a primary care population, even after controlling for medical morbidity, initial primary care utilization, and current medical care utilization [25]. Alcohol use disorders also were related to inconsistent attendance and less complex services for the chronic medical disease, suggesting that medical care may be less accessible for the comorbid population, who may require an increased focus on the complexities of care [26]. Among patients receiving diabetes care in the Department of Veterans Affairs, those who also had SUDs or mental illness had more diabetes complications [27]. Although SUDs have been linked to higher utilization of primary, specialty, and emergency medical services [25], there are few studies of how presence of complex medical disorders affects addiction treatment utilization.

Chronic medical diseases require regular and longitudinal care and self-management for effective treatment [28-30]. When chronic diseases include SUDs, care and treatment of both the medical and addiction disorders may affect access to care and the ability to focus on both conditions. An integrated approach could lead to effective treatment for all conditions [29], yet little is known about how chronic medical disease affects addiction treatment utilization. Chronic disease could cause functional limitations that interfere with an individual's ability to access addiction treatment, or alternatively could increase interaction with the health care system and perhaps addiction treatment. Involvement in addiction treatment could lead to linkage with medical care and effective attention to a chronic medical problem. The objective of this paper is to evaluate the association between the presence of chronic medical disease and recent addiction treatment utilization among adults with substance dependence.

Methods

Data Source and Sample

Data are from baseline interviews with 563 alcohol and/or drug-dependent adults enrolled in the AHEAD (Addiction Health Evaluation And Disease management) study, a randomized controlled trial testing the effectiveness of a chronic disease management program for substance dependence in primary care.

Study recruitment occurred from a free-standing inpatient detoxification unit (74% of enrolled subjects), primary care clinics and the emergency department in the urban medical center where the study was located (10%) and the community by advertising on buses and in newspapers (16%); subjects were recruited between September 2006 and September 2008.

Adults were eligible if they had (1) a diagnosis of current alcohol or drug dependence as assessed by the Composite International Diagnostic Interview Short Form (CIDI-SF) [31]; (2) past 30 day drug use or heavy alcohol use (defined as ≥4 standard drinks for women,

≥5 for men at least twice; or >14 drinks per week for women, >21 drinks per week for men, in an average week in the past month); and (3) were willing to establish or continue primary medical care at the study location. Subjects were excluded if they could not provide contact information for tracking purposes; were not fluent in English or Spanish; had specific plans to leave the area that would preclude in-person research assessments; were pregnant; or had a Mini-Mental State Examination [32] score <21.

Trained research associates administered the standardized baseline research assessment and assured confidentiality. Data collected include sociodemographics, substance use and related problems, anxiety, depression, health questions, and medical and addiction service use. Subjects were compensated for their study participation after completing baseline study procedures.

The Institutional Review Board approved the study. Additional privacy protection was secured by a Certificate of Confidentiality issued by the Department of Health and Human Services.

Chronic Disease and Other Medical Condition Measures

Subjects were asked if a doctor had ever told them they had any of a series of medical conditions. A number of these conditions were included on a chronic disease comorbidity questionnaire validated by Katz et al. [33], which is a self-report version of the medical record-based Charlson Comorbidity Index [34]. The Katz questionnaire includes myocardial infarction, congestive heart failure, peripheral vascular, stroke or other cerebrovascular disease, COPD, ulcer disease, diabetes, renal, connective tissue disease, and other conditions (dementia, liver disease, leukemia, lymphoma, cancer, AIDS). Scoring replicates the Charlson Comorbidity Index, which empirically weighted severity by disease (e.g., cerebrovascular = 1, moderate or severe renal disease = 2, AIDS or metastatic solid tumor = 6), based on 1-year adjusted relative risk for mortality [34]. The comorbidity score is the weighted sum of the specified conditions.

The main independent variable, chronic disease status, is a 3-category variable which classified comorbidity scores as: none; only one condition of lowest severity (Katz = 1); or, more than one condition of lowest severity or one or more condition(s) of higher severity (Katz > 1). Asthma was considered in separate secondary analyses, as an example of a specific symptomatic condition, selected due to its high prevalence in the sample; it is also included in the Katz questionnaire, as part of COPD, with a weight of 1. Asthma is a common comorbidity among urban young to middle-age adults, reflecting the age range common for addiction, whereas other common chronic diseases such as heart disease are less common in young populations. Two other common chronic conditions in this population, hepatitis and hypertension, were prevalent but are often asymptomatic to the patient, and thus were not evaluated in the current analyses.

Additional secondary independent variables were used to evaluate whether broader health status significantly predicted addiction treatment utilization: the single item self-report health status from the SF-12 [35] (excellent/very good/good, fair/poor) and physical health-related functioning and quality of life as determined by the SF-12 Physical Component Score (PCS) (continuous 0-100 scale).

Addiction Treatment Utilization Measures

The primary outcome was any addiction treatment based on self-report, defined as any treatment for addiction (excluding detoxification) in the 3 months prior to study enrollment in any of the following: residential program for alcohol or drug treatment; outpatient full-day/partial hospital program; outpatient with a psychiatrist, other doctor, or other healthcare professional (e.g., counselor); or by taking medication to prevent drinking or drug use. Two secondary outcomes were the counts of outpatient addiction treatment days (at a full-day/partial hospital or with a psychiatrist, other doctor, or other healthcare professional including counselors) and nights in a residential program for alcohol or other drug treatment in the prior 3 months based on self-report at the baseline assessment. These measures were adapted from the COMBINE study Form 90 [36]. Self-report is known to be a reliable way of obtaining treatment utilization information among people with substance use disorders [37,38].

Covariates

Demographic variables included age, gender, and race/ethnicity. Current type of substance dependence is a 3-category variable indicating either alcohol dependence with heavy alcohol use in the past 30 days, any drug dependence with drug use in the past 30 days, or both. Alcohol and drug dependence were determined using the CIDI-SF alcohol and drug modules [31]; heavy alcohol use in the past month was indicated by 4 or more drinks per day or 15 or more drinks per average week for women, and 5 or more drinks per day or 22 or more drinks per week for men. The regression models included other variables that may affect addiction treatment utilization: recruitment from a detoxification program (versus from medical care or the community), criminal justice involvement (any arrest, probation, parole, pretrial release, diversion program or incarceration in the past 3 months), homeless (any time living in a shelter or on the streets in the past 3 months),

unemployed (usual pattern past 3 months), having health insurance (including Medicaid), current smoker, and anxiety in the past week (Beck Anxiety Index [39] scores: low = 0-21, moderate = 22-35, or high = 36 or higher).

Statistical Methods

Secondary data analysis of the AHEAD study baseline data was conducted. Descriptive statistics were obtained and Spearman correlations were used to evaluate potential collinearity between independent variables and covariates. No pair of variables included in the same regression model was highly correlated ($r > 0.40$). Multivariate regression models were used to evaluate the effect of presence of chronic disease on addiction treatment utilization, adjusting for covariates defined above.

The primary outcome of any addiction treatment was modeled using logistic regression, with odds ratios and 95% confidence intervals calculated. The distributions for outpatient addiction treatment days and residential nights, which are count variables, were skewed with many zeros and long tails, thus models assuming normality were inappropriate and overdispersed Poisson regression models were used [40]. The Pearson chi-square correction was used to account for overdispersion. The magnitudes of association between measures of chronic disease and utilization were quantified for the Poisson models using incidence rate ratios (IRRs) with 95% confidence intervals. IRR is interpreted as the ratio of the number of outpatient days (or residential nights) for the exposed group of interest (e.g., any chronic disease) versus the reference group (no chronic disease). The null value of no association for the IRR is equal to 1. An IRR >1 indicates that chronic disease is associated with more treatment utilization.

Multivariate models were fit separately for each dependent variable (any addiction treatment, outpatient days, and residential nights) and for each key independent variable (chronic disease, asthma, health status, and PCS). The primary analyses using the Katz score included 498 subjects with complete data on outcomes, the main independent variable, and covariates; secondary models included the same 498 subjects. Power calculations assumed that 27% of subjects without chronic disease utilized addiction treatment (based on the observed data), giving our study approximately 80% power to detect an odds ratio as small as 1.9 for those with one chronic disease of lowest severity and an odds ratio as small as 2.1 for those with one chronic disease of higher severity or multiple chronic diseases. Thus, power was sufficient for effects that were clinically meaningful. All analyses were conducted using two-sided tests and a significance level of 0.05, using SAS software (version 9.1; SAS Institute, Cary, NC).

Results

Baseline Characteristics

Of 2,018 potential subjects screened, 650 subjects were eligible, and 563 subjects were enrolled. Enrolled subjects were similar to unenrolled subjects in terms of gender, age and race/ethnicity. Baseline sociodemographic characteristics are noted in Table 1, as are prior healthcare and addiction treatment utilization.

Table 1 Baseline Characteristics (n = 563)

	n	%
Age	Mean = 38.2, range 18-67	
Male	409	73
Race		
Black/African American	185	33
White	272	48
Other	106	19
Hispanic or Latino	76	14
Homeless (past 3 mo)	332	59
Unemployed (past 3 mo)	246	44
Has any insurance	446	79
Criminal justice involvement past 3 mo		
None	315	56
Probation	130	23
Pretrial release	63	11
Other/Missing	55	10
Substance dependence		
Alcohol dependent & recent heavy use	98	17
Drug dependent & recent use	150	27
Both	315	56
Substance reported as the major problem		
Alcohol to intoxication	142	25
Heroin	208	37
Cocaine	81	14
Poly-substance or other drugs	132	23
Recruitment site		
Detoxification unit	416	74
Ambulatory/outpatient/ER	59	10
Community/other	88	16
Current smoker	493	88
Beck anxiety score		
Low	220	40
Moderate	164	30
Severe	166	30
Prior healthcare utilization past 3 mo		
Any emergency room	304	54
Any non-addiction outpatient	242	43
Addiction treatment utilization past 3 mo		
Any (excluding detoxification)	157	28
Outpatient days	97	17
Residential nights	70	12

Addiction Treatment

At baseline, 28% of subjects had been in addiction treatment in the prior 3 months, excluding detoxification (Table 1). The 97 patients (17%) who had outpatient treatment in the prior 3 months attended a median of 5 days (interquartile range 3-13). The 70 patients (12%) with residential treatment in the past 3 months attended a median of 23 nights (interquartile range 8-45).

Chronic Disease, Physical Health and Other Medical Conditions

Overall, two thirds of subjects had no chronic disease, 20% had one chronic disease of lower severity, and 14% had one of higher severity or multiple diseases (Table 2). One third of subjects reported their health status to be only poor or fair. The mean Physical Comorbidity Score is 41.7 (SD = 8.4) (the general population norm is 50 [35]).

Subjects reported a wide range of medical conditions (Table 2). Hepatitis was reported by nearly one third of subjects, the most common single condition. Hypertension was reported by 21% of subjects and about 20% reported asthma. Ten to thirteen percent of subjects reported skin infections, pneumonia, anemia or seizures. Five to ten percent reported tachycardia, gastritis, diabetes, COPD, ulcer, or tuberculosis. Few (2-5%) reported rheumatoid arthritis, neuropathy, heart conditions, stroke, HIV, pancreatitis, cancer, poor kidney function or cirrhosis, and remaining conditions were even more rare.

Effect of Chronic Disease on Addiction Treatment Utilization

In multivariate regression analyses (Table 3), no significant effect was detected for the main independent variable of chronic disease status on the odds of attending addiction treatment (adjusted odds ratio [AOR] 0.88 for lower severity vs. no chronic disease, 95% confidence interval (CI): 0.60, 1.28; AOR 1.29 for higher severity vs. no chronic disease, 95% CI: 0.89, 1.88). Similarly, no significant effect of chronic disease status was detected on number of outpatient days or residential nights.

For the primary outcome of any addiction treatment utilization (Model 1), being black was a negative factor, and having health insurance and severe anxiety were positively associated with any addiction treatment. Being female and having severe anxiety were positively associated with number of outpatient days, and recruitment from a detoxification facility was a negative factor (Model 2). Being black was a negative factor for residential nights (Model 3).

Similar results were found for the secondary independent variables of asthma and PCS, with neither significantly associated with any of the addiction treatment measures. Fair or poor health status was negatively associated with any addiction treatment and with outpatient addiction treatment, but not with residential treatment. Table 4 summarizes the results for each of the secondary independent variables (from separate regression models) on each of the outcome variables.

Table 2 Chronic Disease and Other Medical Conditions (n = 563)

	n	%
Summary measures:		
Chronic disease (Katz)		
None	373	67
One, of lower severity only	109	20
One of higher severity or multiple conditions	75	14
Health status (self-reported)		
Excellent	48	9
Very Good/Good	331	59
Fair/Poor	184	33
Physical Comorbidity Score (SF-12)	mean = 41.7, range 17-60	
Has a doctor ever told you that you had:[1]		
Hepatitis	181	32
Hypertension*	120	21
Asthma*	111	20
Skin infections	74	13
Pneumonia	73	13
Anemia	71	13
Seizures, epilepsy or convulsions	60	11
Rapid heart beat or tachycardia	45	8
Gastritis	40	7
Diabetes*	37	7
Emphysema, chronic bronchitis, COPD*	35	6
Ulcer	32	6
Tuberculosis	28	5
Rheumatoid arthritis*	24	4
Peripheral neuropathy	21	4
Heart attack*	19	3
Heart failure	18	3
Stroke*	16	3
HIV*	16	3
Pancreatitis	14	2
Cancer*	13	2
Poor kidney function*	12	2
Cirrhosis	11	2

* Condition included in Katz Comorbidity Index

[1] <2% had blood clots in legs or lungs, septic arthritis, endocarditis, atrial fibrillation, cancer (mouth, throat, larynx, esophagus, stomach), peripheral vascular disease*, AIDS*, lupus*, leukemia* or polycythemia vera*, Alzheimer's or other dementia*, lymphoma*, kidney transplant* or polymylagia rheumatica*

Table 3 Multivariate Regression of Chronic Disease on Addiction Treatment Utilization (N = 498)

	Model 1 Any Addiction Treatment	Model 2 Outpatient Days	Model 3 Residential Nights
Independent Variable	**AOR (95% CI)**	**IRR (95% CI)**	**IRR (95% CI)**
Chronic disease (Katz)			
None	Referent	Referent	Referent
One, of lower severity only	0.88 (0.60, 1.28)	0.86 (0.41, 1.81)	0.68 (0.31, 1.49)
One of higher severity or multiple conditions	1.29 (0.89, 1.88)	1.14 (0.55, 2.38)	0.63 (0.23, 1.69)
Age	0.99 (0.98, 1.01)	1.00 (0.97, 1.04)	0.99 (0.96, 1.03)
Female	1.23 (0.89, 1.70)	1.74 (0.95, 3.20)	0.68 (0.31, 1.48)
Race			
Black	0.57 (0.38, 0.85)**	0.47 (0.21, 1.05)	0.22 (0.08, 0.64)**
Hispanic	0.77 (0.49, 1.20)	1.34 (0.62, 2.89)	0.55 (0.22, 1.36)
Other	0.61 (0.32, 1.15)	0.75 (0.22, 2.57)	0.77 (0.27, 2.24)
White	Referent	Referent	Referent
Substance dependence			
Alcohol dependent & recent heavy use	0.80 (0.52, 1.24)	0.78 (0.33, 1.83)	1.13 (0.50, 2.57)
Drug dependent & recent use	0.86 (0.61, 1.22)	0.87 (0.39, 1.95)	0.48 (0.21, 1.11)
Both	Referent	Referent	Referent
Recruitment site			
Detoxification program	0.82 (0.57, 1.17)	0.37 (0.19, 0.73)**	1.11 (0.48, 2.57)
Other	Referent	Referent	Referent
Criminal justice involvement (past 3 mo)	0.98 (0.73, 1.31)	1.26 (0.70, 2.27)	1.20 (0.66, 2.18)
Homeless (past 3 mo)	1.07 (0.80, 1.43)	0.95 (0.53, 1.72)	1.06 (0.58, 1.95)
Unemployed (past 3 mo)	1.07 (0.79, 1.45)	1.60 (0.79, 3.24)	0.85 (0.47, 1.54)
Health insurance	2.04 (1.24, 3.36)**	3.06 (0.88, 10.68)	2.54 (0.96, 6.71)
Current smoker	0.89 (0.61, 1.30)	0.79 (0.34, 1.81)	1.48 (0.53, 4.11)
Anxiety (Beck)			
Low	Referent	Referent	Referent
Moderate	1.45 (0.99, 2.14)	1.71 (0.80, 3.66)	2.15 (0.97, 4.75)
Severe	1.77 (1.23, 2.53)**	2.12 (1.02, 4.41)*	2.11 (0.94, 4.75)

Model 1 used logistic regression; Models 2 and 3 used Poisson regression.

*p < .05 **p < .01; Wald chi-square tests, df = 1

Discussion

Despite the potential for interference or facilitation of chronic medical disease on utilization of addiction care, no statistically significant effect was found for chronic disease status on addiction treatment utilization in the prior 3 months, in this cohort of alcohol and drug dependent persons, controlling for sociodemographics, type of substance dependence, recruitment site, current

Table 4 Summary of Other Secondary Multivariate Regression Models, with alternate key independent variables (N = 498)

	Model 1 Any Addiction Treatment	Model 2 Outpatient Days	Model 3 Residential Nights
Key Independent Variable	**AOR (95% CI)**	**IRR (95% CI)**	**IRR (95% CI)**
Asthma	1.05 (0.74, 1.49)	1.03 (0.52, 2.03)	1.03 (0.50, 2.12)
Self-reported health status			
Excellent/Very Good/Good	Referent	Referent	Referent
Fair/Poor	0.67 (0.48, 0.94)*	0.40 (0.20, 0.80)**	0.61 (0.31, 1.21)
Physical Component Score (per 1 point increase)	1.01 (0.99, 1.03)	1.00 (0.97, 1.04)	1.01 (0.97, 1.04)

Model 1 used logistic regression; Models 2 and 3 used Poisson regression. Models duplicate those in Table 3 (controlling for age, gender, race, substance dependence, recruitment from detox, criminal justice involvement, homeless, unemployed, health insurance, smoker, anxiety), solely replacing the chronic disease variable with the secondary variables above.

*p < .05 **p < .01; Wald chi-square tests, df = 1

smoking, and severity of recent anxiety. Nor was addiction treatment associated with asthma or physical function and quality of life. These findings held for any addiction treatment and for quantity of treatment services received. Only perceived health status, which is a subjective measure, was associated with less addiction treatment utilization. It is encouraging to note that chronic disease itself may not be a barrier to addiction treatment.

On the other hand, it also appears that chronic medical disease did not increase addiction treatment participation in this cohort. People with chronic disease are often connected to the medical treatment system, but the acute focus of episodic care may not extend to providing assistance or motivation for these patients to access addiction treatment. This may be a missed opportunity for medical providers to encourage their patients to seek addiction treatment given that the self-reported perception of health status was frequently only fair or poor. The lack of a significant positive association between chronic disease and addiction treatment utilization suggests that facilitated access to addiction care is not routinely occurring. Such "reachable moments" [41] for the individual with both chronic disease and addiction are perhaps most likely when formalized linkages between primary care and addiction treatment exist or if the primary care doctor is willing to address addiction with the patient [18,42]. One way to increase these reachable moments is to encourage screening, brief intervention, and referral to treatment approaches in the primary care setting. Opportunities for referral to addiction treatment should increase as the number of medical visits increases. An integrated, longitudinal approach to managing both chronic medical disease and substance dependence is likely to be most beneficial [13,43,44].

It is worthwhile to note that no chronic diseases were reported by two thirds of this sample of substance dependent individuals, somewhat better than an estimate of 53% with no chronic conditions in a comparable population 10 years earlier [22]. The lower estimate here may be due to recruitment beyond detoxification settings; subjects recruited from the community may be less medically ill than those in detoxification facilities. Most of the chronic diseases noted here had prevalence within the range of national and state estimates [45,46]. The asthma prevalence here appears slightly higher than the state average, but this likely reflects the higher rates often found in inner city and African-American populations [47] and the high prevalence of current smoking in this sample. The below average Physical Comorbidity Score in this population is comparable to similar populations 10-15 years earlier [22,48].

Although we did not detect an association between chronic medical variables and treatment utilization, several other variables were significantly related to treatment utilization, including several sociodemographic variables. Health insurance was positively associated with any addiction treatment utilization, demonstrating an expected increase in access to treatment, but did not reach significance for the count variables. Recruitment from detoxification was negatively associated with outpatient days, highlighting the severity of addiction treatment needs for this group, and the expected lack of treatment preceding detoxification entry. Psychiatric comorbidity (e.g., anxiety) was positively associated with addiction treatment utilization reflecting other findings in the literature [49], although this past-week measure was collected for most individuals soon after detoxification, so may reflect short-term anxiety symptoms rather than a chronic disorder.

The lack of findings concerning our primary hypothesis may simply be the reflection of a complex relationship between chronic disease and addiction, with many competing forces. Several opposing effects from chronic disease may have occurred, which could collectively show little or no effect on addiction treatment utilization. First, individuals with chronic disease may be functionally less able to access and participate in addiction treatment, reducing the likelihood of treatment utilization. If physical function were an important barrier, we should have seen some effect from the PCS, which did not occur, but it might be important to consider specific functional limitations. By considering asthma separately, thus removing the variability across conditions, we might have highlighted the functional effects, but that also did not occur. It is also possible that the severity of addiction, as indicated by most individuals in the sample recruited from detoxification, overshadowed any functional limitations as they sought treatment; a desire to focus on the most pressing problem has been highlighted in some literature [9]. Second, for a variety of reasons, addiction treatment facilities may be less likely to accept individuals with medical conditions, who may require medications or have more acute medical needs. This study could not address that question. Third, chronic disease may interfere with the quality of treatment participation; number of visits could be construed as a proxy for this, but if chronic disease affects self-care management this issue may go beyond what could be considered here. Fourth, 3 months is a relatively short time period, so may have masked any differences that would appear with longer history of treatment utilization. However, the focus here is on relative differences in utilization, so the 3 month period should have been sufficient to compare groups. Fifth, and alternatively, individuals with chronic disease may be more likely to access treatment due to their likely existing linkages with the healthcare system. A study of women with

trauma and mental or substance use disorders found that greater disability was associated with more outpatient counseling, but not residential, group counseling or peer treatment [50]. While referral from the healthcare system is not a common pathway to addiction treatment [51] the finding that physicians are more likely to know about substance abuse if the patient has an episodic or chronic medical condition [11] suggests that this pathway may be more likely for these patients.

The characteristics of the study sample may have tempered the potential relationships between chronic disease and addiction treatment utilization. In a relatively young sample such as this, chronic disease is less likely, and related problems are likely to be less severe than in an older population. However, addiction itself may increase such problems even in a young sample [22]. The subjects in this study are largely reflective of individuals in addiction treatment, and these results should apply to similar cohorts. Relatively few individuals in this sample had severe chronic disease or multiple conditions, thus findings may be different in an analysis of only individuals with chronic disease. The implications of recruiting much of the sample during a detoxification stay should be considered, where individuals entering detoxification may be unlikely to have been in addiction treatment during a period of substance use severe enough to warrant detoxification. These findings also represent a sample from a single site, so may be less generalizable than recruitment from multiple sites.

A limitation is that the study may have been inadequately powered to detect differences of the observed magnitude. Power calculations estimated 80% power to detect an odds ratio as small as 2.1 for those with higher severity chronic disease, thus it is likely that the study was not adequately powered to detect the small observed magnitude of association of 1.3 (for higher severity). However, since the observed degree of association may still be clinically meaningful, the possibility that individuals with higher comorbidity are more likely to utilize addiction treatment should be further studied in a larger sample to see if a small but clinically important effect is detectable. But the effect of greater concern, interference with receiving addiction treatment due to a chronic disease, is unlikely based on our findings. Last, the current analysis relied on retrospective data from a sample of individuals, most of whom had recently completed detoxification. A prospective or qualitative study would be better able to track the likely causal issues related to chronic disease and addiction treatment access.

To our knowledge, this is the first analysis of the role of chronic disease in addiction treatment utilization. Current household studies on barriers to addiction treatment, such as SAMHSA's National Survey on Drug Use and Health (NSDUH), do not consider medical comorbidities, so are unable to shed light on these underlying questions. Further research in this area is warranted to better understand these complex issues. The individuals in this study were, by definition, willing to receive health care. Other populations may differ. Also, samples with a large number of individuals with chronic disease would allow a focus on more specific chronic diseases, such as diabetes or heart disease, and would allow a better consideration of other psychiatric comorbidity. In the meantime, we highlight the conclusion from these data that chronic disease is not a significant barrier to addiction treatment utilization. While we found no significant association between chronic medical disease and addiction treatment utilization, these issues are complicated and we suggest prospective and qualitative studies to further explore these complexities.

Acknowledgements

Preliminary results of this study were presented at the *Integrating Services & Research for Co-Occurring Conditions* conference, Bethesda, MD, March 2009, and the *Association for Medical Education and Research in Substance Abuse* conference, Bethesda MD, November 2009. This study was supported by the National Institute on Alcohol Abuse and Alcoholism (grant R01 AA010870) and the National Institute on Drug Abuse (grant R01 DA010019). This work was also supported, in part, by the Boston University General Clinical Research Unit which is part of the Clinical and Translational Science Institute (National Center for Research Resources (NCRR) Grant Number UL1RR025771), and the Brandeis/Harvard NIDA Research Center (P50 DA010233). JS receives support from NIAAA K24 AA015674. The authors thank Julie Witas for project management, and the AHEAD study research team and subjects.

Author details

[1]Institute for Behavioral Health, Heller School for Social Policy and Management, Brandeis University, 415 South Street, MS035, Waltham, MA 02454, USA. [2]Boston University School of Public Health, 801 Massachusetts Avenue, Boston, MA 02118, USA. [3]Boston University School of Medicine, 801 Massachusetts Avenue, Boston, MA 02118, USA. [4]Boston Medical Center, 801 Massachusetts Avenue, Boston, MA 02118, USA.

Authors' contributions

All authors contributed to the design, interpretation of results, and the manuscript preparation and have approved the final manuscript. In addition, significant individual contributions are as follows. SR conceived of the study, designed the analytic plan, and led the writing, interpretation of results, and revisions of the manuscript. MJL contributed to the study scope, analytic plan, and interpretation of results. DMC provided statistical expertise, and contributed to the study scope, analytic plan, and interpretation of results. DA conducted the statistical analysis. RS and JS lead the parent study from which the data were obtained, brought expertise regarding the data and the study population, and contributed to the study scope, analytic plan, and interpretation of results.

Competing interests

The authors declare that they have no competing interests.

Received: 9 June 2011 Accepted: 18 October 2011
Published: 18 October 2011

References

1. National Health Expenditure Data, Historical. [http://www.cms.hhs.gov/NationalHealthExpendData/02_NationalHealthAccountsHistorical.asp#TopOfPage].

2. **Chronic Disease Overview.** [http://www.cdc.gov/NCCdphp/overview.htm].
3. Clark RE, Samnaliev M, McGovern MP: **Impact of substance disorders on medical expenditures for medicaid beneficiaries with behavioral health disorders.** *Psychiatr Serv* 2009, **60**:35-42.
4. Mark TL, Levit KR, Coffey RM, McKusick DR, Harwood HJ, King EC, Bouchery E, Genuardi JS, Vandivort-Warren R, Buck JA, Ryan K: **National Expenditures for Mental Health Services and Substance Abuse Treatment, 1993-2003.** Rockville, MD: Substance Abuse and Mental Health Services Administration; 2007.
5. Office of National Drug Control Policy: **The Economic Costs of Drug Abuse in the United States, 1992-2002.** Washington, D.C.: Executive Office of the President; 2004.
6. Substance Abuse and Mental Health Services Administration: **Results from the 2008 National Survey on Drug Use and Health: National Findings.** Rockville, MD: Office of Applied Studies; 2009.
7. Paez KA, Zhao L, Hwang W: **Rising out-of-pocket spending for chronic conditions: a ten-year trend.** *Health Aff (Millwood)* 2009, **28**:15-25.
8. Decker SL, Schappert SM, Sisk JE: **Use of medical care for chronic conditions.** *Health Aff (Millwood)* 2009, **28**:26-35.
9. Redelmeier DA, Tan SH, Booth GL: **The treatment of unrelated disorders in patients with chronic medical diseases.** *N Engl J Med* 1998, **338**:1516-1520.
10. Borowsky SJ, Rubenstein LV, Meredith LS, Camp P, Jackson-Triche M, Wells KB: **Who is at risk of nondetection of mental health problems in primary care?** *J Gen Intern Med* 2000, **15**:381-388.
11. Saitz R, Mulvey KP, Plough A, Samet JH: **Physician unawareness of serious substance abuse.** *Am J Drug Alcohol Abuse* 1997, **23**:343-354.
12. Graham AW, Schultz TK, Mayo-Smith MF, Ries RK, Wilford BB: *Principles of Addiction Medicine.* 3 edition. Chevy Chase, MD: American Society of Addiction Medicine; 2003.
13. Institute of Medicine: *Improving the Quality of Health Care for Mental and Substance-Use Conditions* Washington, D.C.: National Academies Press; 2006.
14. Banerjea R, Sambamoorthi U, Smelson D, Pogach LM: **Expenditures in mental illness and substance use disorders among veteran clinic users with diabetes.** *J Behav Health Serv Res* 2008, **35**:290-303.
15. Garnick DW, Hendricks AM, Comstock C, Horgan C: **Do individuals with substance abuse diagnoses incur higher charges than individuals with other chronic conditions?** *J Subst Abuse Treat* 1997, **14**:457-465.
16. Thomas MR, Waxmonsky JA, Gabow PA, Flanders-McGinnis G, Socherman R, Rost K: **Prevalence of psychiatric disorders and costs of care among adult enrollees in a Medicaid HMO.** *Psychiatr Serv* 2005, **56**:1394-1401.
17. Wadland WC, Ferenchick GS: **Medical comorbidity in addictive disorders.** *Psychiatr Clin North Am* 2004, **27**:675-687.
18. Mertens JR, Lu YW, Parthasarathy S, Moore C, Weisner CM: **Medical and psychiatric conditions of alcohol and drug treatment patients in an HMO: comparison with matched controls.** *Arch Intern Med* 2003, **163**:2511-2517.
19. Larson MJ, Paasche-Orlow M, Cheng DM, Lloyd-Travaglini C, Saitz R, Samet JH: **Persistent pain is associated with substance use after detoxification: a prospective cohort analysis.** *Addiction* 2007, **102**:752-760.
20. Dickey B, Normand SL, Weiss RD, Drake RE, Azeni H: **Medical morbidity, mental illness, and substance use disorders.** *Psychiatr Serv* 2002, **53**:861-867.
21. Fox K, Merrill JC, Chang HH, Califano JA Jr: **Estimating the costs of substance abuse to the Medicaid hospital care program.** *Am J Public Health* 1995, **85**:48-54.
22. De Alba I, Samet JH, Saitz R: **Burden of medical illness in drug- and alcohol-dependent persons without primary care.** *Am J Addict* 2004, **13**:33-45.
23. Mee-Lee D, Shulman GD, Fishman M, Gastfriend DR, Griffith JH, (Eds.): *ASAM Patient Placement Criteria for the Treatment of Substance-Related Disorders, 2nd edition-revised* Chevy Chase, MD: American Society of Addiction Medicine; 2001.
24. Schulz JE: **The integration of medical management with recovery.** *J Psychoactive Drugs* 1997, **29**:233-237.
25. Ford JD, Trestman RL, Steinberg K, Tennen H, Allen S: **Prospective association of anxiety, depressive, and addictive disorders with high utilization of primary, specialty and emergency medical care.** *Soc Sci Med* 2004, **58**:2145-2148.
26. Ford JD, Trestman RL, Tennen H, Allen S: **Relationship of anxiety, depression and alcohol use disorders to persistent high utilization and potentially problematic under-utilization of primary medical care.** *Soc Sci Med* 2005, **61**:1618-1625.
27. Banerjea R, Sambamoorthi U, Smelson D, Pogach LM: **Chronic illness with complexities: mental illness and substance use among Veteran clinic users with diabetes.** *Am J Drug Alcohol Abuse* 2007, **33**:807-821.
28. Merrick EL, Horgan CM, Garnick DW, Hodgkin D, Morley M: **Health plans' disease management programs: extending across the medical and behavioral health spectrum?** *J Ambul Care Manage* 2008, **31**:342-353.
29. Saitz R, Larson MJ, LaBelle C, Richardson J, Samet JH: **The case for chronic disease management for addiction.** *J Addict Med* 2008, **2**:55-65.
30. Wagner EH, Austin BT, Von Korff M: **Organizing care for patients with chronic illness.** *Milbank Q* 1996, **74**:511-544.
31. Kessler RC, Andrews G, Mroczek D, Ustun B, Wittchen HU: **The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF).** *Int J Methods Psychiatr Res* 1998, **7**:171-185.
32. Smith KL, Horton NJ, Saitz R, Samet JH: **The use of the mini-mental state examination in recruitment for substance abuse research studies.** *Drug Alcohol Depend* 2006, **82**:231-237.
33. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW: **Can comorbidity be measured by questionnaire rather than medical record review?** *Med Care* 1996, **34**:73-84.
34. Charlson ME, Pompei P, Ales KL, MacKenzie CR: **A new method of classifying prognostic comorbidity in longitudinal studies: development and validation.** *J Chronic Dis* 1987, **40**:373-383.
35. Ware J Jr, Kosinski M, Keller SD: **A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity.** *Med Care* 1996, **34**:220-233.
36. Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM, Gastfriend DR, Hosking JD, Johnson BA, LoCastro JS, *et al*: **Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE study: a randomized controlled trial.** *JAMA* 2006, **295**:2003-2017.
37. Beebe TJ, McRae JA Jr, Barnes SA: **A comparison of self-reported use of behavioral health services with Medicaid agency records in Minnesota.** *Psychiatr Serv* 2006, **57**:1652-1654.
38. Killeen TK, Brady KT, Gold PB, Tyson C, Simpson KN: **Comparison of self-report versus agency records of service utilization in a community sample of individuals with alcohol use disorders.** *Drug Alcohol Depend* 2004, **73**:141-147.
39. Beck AT, Epstein N, Brown G, Steer RA: **An inventory for measuring clinical anxiety: psychometric properties.** *J Consult Clin Psychol* 1988, **56**:893-897.
40. McCullagh P, Nelder JA: *Generalized Linear Models.* 2 edition. London: Chapman and Hall; 1989.
41. Shanahan CW, Beers D, Alford DP, Brigandi E, Samet JH: **A transitional opioid program to engage hospitalized drug users.** *J Gen Intern Med* 2010, **25**:803-808.
42. Samet JH, Friedmann P, Saitz R: **Benefits of linking primary medical care and substance abuse services: patient, provider, and societal perspectives.** *Arch Intern Med* 2001, **161**:85-91.
43. Weisner C, Mertens J, Parthasarathy S, Moore C, Lu Y: **Integrating primary medical care with addiction treatment: a randomized controlled trial.** *JAMA* 2001, **286**:1715-1723.
44. Weisner C, Ray GT, Mertens JR, Satre DD, Moore C: **Short-term alcohol and drug treatment outcomes predict long-term outcome.** *Drug Alcohol Depend* 2003, **71**:281-294.
45. National Center for Health Statistics: **Summary Health Statistics for U.S. Adults: National Health Interview Survey, 2008 (provisional report).** *Vital and Health Statistics, Series 10, Number 242* Hyattsville, MD: Centers for Disease Control and Prevention; 2009.
46. Massachusetts Department of Public Health: **A Profile of Health Among Massachusetts Adults, 2008: Results from the Behavioral Risk Factor Surveillance System.** Boston, MA: Massachusetts Department of Public Health; 2009.
47. Gold DR, Wright R: **Population disparities in asthma.** *Annu Rev Public Health* 2005, **26**:89-113.
48. Stein MD, Mulvey KP, Plough A, Samet JH: **The functioning and well being of persons who seek treatment for drug and alcohol use.** *J Subst Abuse* 1998, **10**:75-84.
49. Ilgen MA, Price AM, Burnett-Zeigler I, Perron B, Islam K, Bohnert AS, Zivin K: **Longitudinal predictors of addictions treatment utilization in treatment-**

naive adults with alcohol use disorders. *Drug Alcohol Depend* 2011, **113**:215-221.

50. Larson MJ, Miller L, Becker M, Richardson E, Kammerer N, Thom J, Gampel J, Savage A: **Physical health burdens of women with trauma histories and co-occurring substance abuse and mental disorders.** *J Behav Health Serv Res* 2005, **32**:128-140.
51. Substance Abuse and Mental Health Services Administration: **Treatment Episode Data Set (TEDS) - Highlights 2007. National Admissions to Substance Abuse Treatment Services.** Rockville, MD: Office of Applied Studies; 2009.

doi:10.1186/1747-597X-6-28
Cite this article as: Reif *et al.*: Chronic disease and recent addiction treatment utilization among alcohol and drug dependent adults. *Substance Abuse Treatment, Prevention, and Policy* 2011 **6**:28.

Circulation
Cardiovascular Quality and Outcomes
JOURNAL OF THE AMERICAN HEART ASSOCIATION
American Heart Association
Learn and Live

The Business Case for Quality Improvement : Oral Anticoagulation for Atrial Fibrillation

Adam J. Rose, Dan R. Berlowitz, Arlene S. Ash, Al Ozonoff, Elaine M. Hylek and Jeremy D. Goldhaber-Fiebert

Circ Cardiovasc Qual Outcomes 2011;4;416-424; originally published online June 28, 2011;
DOI: 10.1161/CIRCOUTCOMES.111.960591
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514
 Print ISSN: 1941-7705. Online ISSN: 1941-7713

The Business Case for Quality Improvement
Oral Anticoagulation for Atrial Fibrillation

Adam J. Rose, MD, MSc, FACP; Dan R. Berlowitz, MD, MPH; Arlene S. Ash, PhD; Al Ozonoff, PhD; Elaine M. Hylek, MD, MPH; Jeremy D. Goldhaber-Fiebert, PhD

Background—The potential to save money within a short time frame provides a more compelling "business case" for quality improvement than merely demonstrating cost-effectiveness. Our objective was to demonstrate the potential for cost savings from improved control in patients anticoagulated for atrial fibrillation.

Methods and Results—Our population consisted of 67 077 Veterans Health Administration patients anticoagulated for atrial fibrillation between October 1, 2006, and September 30, 2008. We simulated the number of adverse events and their associated costs and utilities, both before and after various degrees of improvement in percent time in therapeutic range (TTR). The simulation had a 2-year time horizon, and costs were calculated from the perspective of the payer. In the base-case analysis, improving TTR by 5% prevented 1114 adverse events, including 662 deaths; it gained 863 quality-adjusted life-years and saved $15.9 million compared with the status quo, not accounting for the cost of the quality improvement program. Improving TTR by 10% prevented 2087 events, gained 1606 quality-adjusted life-years, and saved $29.7 million. In sensitivity analyses, costs were most sensitive to the estimated risk of stroke and the expected stroke reduction from improved TTR. Utilities were most sensitive to the estimated risk of death and the expected mortality benefit from improved TTR.

Conclusions—A quality improvement program to improve anticoagulation control probably would be cost-saving for the payer, even if it were only modestly effective in improving control and even without considering the value of improved health. This study demonstrates how to make a business case for a quality improvement initiative. **(*Circ Cardiovasc Qual Outcomes*. 2011;4:416-424.)**

Key Words: anticoagulants ■ atrial fibrillation ■ patient simulation ■ quality improvement

Atrial fibrillation (AF) is a common condition and a leading cause of ischemic stroke.[1,2] The benefit of long-term anticoagulation with warfarin to prevent ischemic stroke in patients with AF is well established.[3–6] However, the anticoagulation that is actually provided leaves much room for improvement.[7] Quality of anticoagulation can be measured by percent time in the therapeutic range (TTR).[8,9] Higher TTR has been linked with lower rates of ischemic stroke, major hemorrhage, and death.[10,11] Although improving TTR would have benefits, payers may be hesitant to invest the time, effort, and resources required for a quality improvement (QI) program. In particular, payers are generally less interested in questions of long-term cost-effectiveness, focusing instead on traditional business considerations of maximizing short-term profitability. Because the Veterans Health Administration (VA) is considering a QI program to increase TTR, we undertook to study whether a "business case" can be made for such a program, that is, whether it has the potential to save money in the short term.

We therefore used our Veterans AffaiRs Study to Improve Anticoagulation (VARIA) data base to identify 67 077 patients with AF receiving anticoagulation from the VA over a 2-year period. We simulated the number of adverse events that would be prevented in this population through improved TTR as well as the resulting cost savings and utility gains. We hypothesized that potential cost savings to the payer (the VA), mostly from preventing ischemic strokes, would constitute a compelling business case for this QI program.

Received January 11, 2011; accepted April 13, 2011.

From the Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, Bedford, MA (A.J.R., D.R.B., A.O.); the Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA (A.J.R., D.R.B., E.M.H.); the Department of Health Policy and Management, Boston University School of Public Health, Boston, MA (D.R.B.); the Department of Quantitative Health Sciences, Division of Biostatistics and Health Services Research, University of Massachusetts School of Medicine, Worcester, MA (A.S.A.); the Biostatistics Section, Boston Children's Hospital, Boston, MA (A.O.); and the Centers for Health Policy and Primary Care and Outcomes Research, Stanford University School of Medicine, Stanford, CA (J.D.G.-F.).

The online-only Data Supplement is available at http://circoutcomes.ahajournals.org/cgi/content/full/CIRCOUTCOMES.111.960591/DC1.

Correspondence to Adam J. Rose, MD, MSc, FACP, Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, 200 Springs Rd, Bldg 70, Bedford, MA 01730. E-mail adamrose@bu.edu

Circ Cardiovasc Qual Outcomes **is available at http://circoutcomes.ahajournals.org** **DOI: 10.1161/CIRCOUTCOMES.111.960591**

WHAT IS KNOWN

- Improved control of oral anticoagulation could prevent many adverse events, which would improve patient health and could potentially produce cost savings for payers.
- Cost savings from adverse events averted could at least partly defray the costs of a quality improvement program to improve anticoagulation control.

WHAT THE STUDY ADDS

- Using data from a real cohort of >60 000 patients anticoagulated for atrial fibrillation in the Veterans Health Administration, we simulated the number of adverse events that might be prevented by slight, moderate, and large improvements in anticoagulation control.
- Even assuming a small improvement in control and even with other conservative assumptions, the short-term cost savings from averted events would likely be much larger than the historical cost of most quality improvement programs.
- Because improved anticoagulation control has the potential to produce short-term savings from the payer's perspective while improving patient health, pursuing this goal should be a top priority for the Veterans Health Administration and other systems of care.

Methods

Inclusion Criteria and Calculation of Percent Time in Range (TTR)

The data base for this study has been described previously.[9,12] VARIA included all patients deemed to be receiving oral anticoagulation therapy from the VA between October 1, 2006, and September 30, 2008, according to the criteria described below. The study was approved by the Institutional Review Board of the Bedford VA Medical Center.

There are 128 sites of care in the VA system. We included 100 of these sites and excluded the remaining 28 because their international normalized ratio (INR) values were not captured reliably in VA data bases. For the current study, we limited our data base to patients who were experienced users of warfarin, that is, who had used it for at least 6 months. We defined each patient's date of warfarin initiation as the first INR value >1.2 or the first outpatient warfarin fill, whichever came first. It would be extremely unusual for a patient to record an INR value >1.2 unless he or she had taken warfarin. We then stratified the sample into inception time (the first 6 months of warfarin therapy for each patient) and experienced time (any time thereafter).

We included INR tests within the VA when patients were "on warfarin": that is, when a patient was either (1) "in possession" of warfarin or (2) having INR tests every 42 days. We defined the period of warfarin possession as the duration of the most recent VA prescription for warfarin, plus 30 days. We calculated TTR using the Rosendaal method,[13] which uses linear interpolation to calculate the percentage of time during which the INR was between 2 and 3.[13]

Estimates of Adverse Event Risks and Their Relationship to Clinical Characteristics

We focused on 3 adverse events: ischemic stroke, major hemorrhage, and death. We relied on estimates from randomized, controlled trials and observational studies, published in English, that reported all of the following: the anticoagulation control achieved by patients receiving adjusted-dose warfarin; sufficient information to compute rates for at least 1 category of adverse events (ie, number of events and person-time of observation); and information on the proportion of patients with congestive heart failure, hypertension, age >75 years, diabetes mellitus, and prior stroke or transient ischemic attack.

We identified studies using the literature review conducted by the American College of Chest Physicians Guidelines Group in their most recent guidelines for stroke prevention in AF.[5] This review included all randomized and observational studies of patients receiving adjusted-dose warfarin for stroke prevention in AF from 1950 through 2008. We characterized patients' risk of adverse events using the CHADS2 score.[14] CHADS2 stands for Congestive heart failure, Hypertension, Age >75 years, Diabetes, and prior Stroke. Patients are assigned 1 point for each of the first 4 risk factors, and 2 points for prior stroke; scores vary from 0 to 6. CHADS2 scores are predictive of the risk of ischemic stroke[14] as well as major hemorrhage[15–17] and are widely used by clinicians and researchers to summarize stroke risk in AF.

The included studies are listed in Table 1. In total, 18 studies reported rates of ischemic stroke and major hemorrhage[18–35]; of those, 14 studies also reported all-cause mortality.[18–31] For each study, we computed the average CHADS2 score of the included patients.[14] We then used Poisson regression to model the annual rate of each adverse event type as a function of study average CHADS2 score. We assumed that these risks increase exponentially as the CHADS2 score increases, as was shown in the original CHADS2 study.[14] We used the CHADS2 score not only in its original sense (as a predictor of ischemic stroke risk) but also to predict the risk of other adverse events (ie, major hemorrhage and all-cause mortality). The supposition that patients with higher CHADS2 scores have higher rates of all-cause mortality is reasonable but has not been empirically demonstrated. When we subsequently used these models to predict the annual rate of events for individual patients, we substituted each patient's actual CHADS2 score for the study average CHADS2 score. The parameter estimates for these Poisson regressions are shown in Table 2. The pseudo R^2 values for meta-regressions of stroke, major hemorrhage, and death were 0.16, 0.19, and 0.19, respectively.

Estimates of the Effects of TTR on Adverse Event Risks

We then estimated the relationships between TTR and the rates of our 3 adverse events, using analysis from White et al.[11] Although White reported the relationship between TTR and outcomes using tertiles, we used additional information from another study that analyzed the same data sets[36] to produce a continuous plot for event rates as a function of TTR (Figure 1). To model the predicted event rates for each patient, we predicted each event rate for each patient based on his/her CHADS2 score. We then adjusted this rate on the basis of the patient's TTR.

Simulating the Adverse Events for the VA Population

We constructed a simulation model to predict the likelihood of each type of adverse event given current TTR levels (the "status quo") over a 2-year period and used it to explore how improvements in TTR levels would reduce adverse events. First, we predicted the 2-year event rates for each patient on the basis of his or her CHADS2 score and TTR level, as described above, then summed these event rates to describe the overall rate of any event occurring. We converted this summed rate to a daily probability assuming a constant hazard (exponential model). We then drew random numbers to determine whether any event had occurred each day, and, if yes, then of what type. We assumed that 16% of major hemorrhages are intracranial, as per O'Brien et al.[37] We right-censored patients who did not have an event within the 2-year analysis window. We repeated the simulation 1000 times to generate 95% confidence intervals.

Table 1. Randomized Trials and Observational Studies Included in Our Meta-Regression

	Follow-Up and Anticoagulation Control			Risk Factors for Adverse Events, CHADS2 Score Components					Outcomes		
Randomized, Controlled Trials	Patient-Years	n	Mean TTR	Age ≥75 Years*	DM	HTN	Prior CVA or TIA	CHF	Stroke	Major Bleed	Death
AFASAK-1[34]	413	335	73	39	7	32	6	50	2.5	0.5	. . .
BAATAF[18]	487	212	83	22	14	51	3	24	0.2	0.4	2.3
SPAF 1[19]	260	210	71	11	12	49	8	14	1.5	1.5	2.3
CAFA[25]	237	187	44	23	14	43	3	24	2.1	0.8	5.1
SPINAF[26]	456	260	56	13	17	55	10	31	0.7	1.5	3.5
EAFT[20]	507	225	59	28	12	43	100	8	3.9	2.6	8.1
SPAF 2 (age ≤75 y)[21]	1099	358	75	0	17	53	5	17	0.7	1.7	3.3
SPAF 2 (age >75 y)[21]	394	197	72	100	13	52	5	26	1.5	4.2	6.6
AFASAK 2[27]	355	170	73	40	. . .	47	8	. . .	2.2	1.1	4.8
BAFTA[33]	1318	488	67	100	14	53	13	20	1.0	1.9	. . .
SPAF 3[22]	581	523	61	34	20	67	36	45	1.4	2.1	6.0
PATAF[28]	401	131	48	25	25	46	0	1	0.5	0.2	3.0
Pengo et al[31]	181	153	70	43	14	63	0	15	0	2.2	3.3
SIFA[29]	454	454	84	36	15	55	100	33	3.1	1.8	7.0
SPORTIF III[30]	2440	1703	66	28	24	72	24	34	1.8	2.0	3.2
SPORTIF V[23]	3212	1962	68	35	25	81	18	40	0.9	2.9	3.8
ACTIVE W[24]	4315	3371	64	31	21	82	15	31	0.9	2.2	3.7
Observational											
ACTION[35]	2892	3396	67	46	17	48	11	23	1.0	1.9	. . .
ATRIA[32]	12958	7445	63	35	18	52	11	33	1.2	1.5	. . .

DM indicates diabetes mellitus; HTN, hypertension; CVA, cerebrovascular accident; TIA, transient ischemic attack; and CHF, congestive heart failure.

Patient characteristics are given in percentages, and rates are given in events per 100 patient-years.

*For some studies, the average age was reported along with the standard deviation, but the proportion of patients age >75 years was not directly reported. For these studies, proportion with age ≥75 years was estimated by assuming that age was normally distributed.

Simulating the Benefits of Improvements in TTR for the VA Population

At this point, each patient was defined by 2 parameters that determined his risk of adverse events: CHADS2 score and TTR. We then altered the status quo by assuming that a QI program had increased TTR, ranging from a 2.5% absolute improvement in TTR to a 20% absolute improvement. Our estimate for the approximate magnitude of this change is based on a meta-analysis, which showed that dedicated anticoagulation clinics achieve a TTR 8.9% higher than management by individual clinicians.[7] It is reasonable to expect that an intervention to improve VA anticoagulation clinics could improve TTR by a similar margin. We assumed that all patients in the data set would have an equal improvement in TTR. Some patients began with a very high TTR at baseline; if the incremental improvement had exceeded a TTR of 100%, some of the postulated improvement was "lost."

Costs and Utilities

For cost and utility estimates, we relied on published literature (Table 3), especially a cost-effectiveness analysis published by O'Brien and Gage.[37] We calculated the cost savings and health benefits of different degrees of improvement in TTR, compared with the status quo (ie, baseline TTR), based on the number of events averted. We included costs for ongoing warfarin therapy and for adverse events that might be prevented by or caused by warfarin therapy. Costs were expressed in 2008 dollars and were inflation-adjusted when their original sources reported costs from other years.[38,39] Utilities were expressed in terms of quality-adjusted life-years (QALYs), a measure that combines longevity and morbidity.[39] These utilities were derived from a population of patients with AF who are similar to our study population in most respects. Given the short time horizon of the analysis (2 years), we report undiscounted costs and QALYs.

Sensitivity/Uncertainty Analyses

We performed 1-way sensitivity analyses on all inputs to the model, including the relationship between CHADS2 scores and event rates, the relationship between TTR and event rates, costs, and utilities. We also performed 2-way sensitivity analyses using pairs of variables with large effect sizes in the 1-way sensitivity analyses.

Although White and others have reported considerable mortality benefits from improved TTR,[11,40,41] we suspect that these are overestimates of the effect size. Although it is logical that deaths caused by ischemic stroke or major hemorrhage would be prevented, we suspect that these estimates are inflated by the inclusion of other causes of death that would not be prevented by improved anticoagulation control. The reason for this would be that sicker patients have lower TTR[12] and higher rates of all-cause mortality, some of which is unrelated to excessive or insufficient anticoagulation. Therefore, we ran an additional set of analyses under the extreme assumption that improved TTR has no impact on all-cause mortality—not because we think this is likely, but simply to explore the fullest possible impact of this input on our results.

Finally, we performed a probabilistic sensitivity analysis (via Monte Carlo simulation), simultaneously varying all of our parameters on the basis of their uncertainty distributions, described by their plausible ranges (Table 3) for 1000 iterations. For the probabilistic sensitivity analyses, we assumed beta distributions for utilities, triangular distributions for costs, and normal distributions for the

Table 2. Results of Poisson Meta-Regressions for Rates of Ischemic Stroke, Major Hemorrhage, and All-Cause Mortality, Based on CHADS2 Scores (Events Per Patient-Year)

	β	95% CI
Ischemic stroke: 18 studies; 393 events; 32 605 patient-years		
CHADS2 score	0.634	0.373 to 0.894
Study is an RCT	−0.165	−0.402 to 0.073
Constant	−5.482	−5.922 to −5.042
Major hemorrhage: 18 studies; 611 events; 32 605 patient-years		
CHADS2 score	0.379	0.175 to 0.584
Study is an RCT	0.151	−0.033 to 0.335
Constant	−4.746	−5.094 to −4.397
All-cause mortality: 14 studies, 601 events, 15 024 patient-years		
CHADS2 score	0.380	0.228 to 0.532
Constant	−3.973	−4.292 to −3.653

RCT indicates randomized, controlled trial.

Observational studies did not report on all-cause mortality; therefore, the variable for RCT was dropped.

relationships between CHADS2 or TTR and adverse event rates. All analyses were undertaken with Stata/SE 10.1 (StataCorp, College Station, TX). Drs Rose and Goldhaber-Fiebert had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Results

Patients

Our population consisted of 67 077 patients anticoagulated for AF between October 1, 2006, and September 30, 2008 (Table 4). Patients were overwhelmingly male (99%), with a mean age of 72.7 years. Patients had considerable comorbid illness: 41% had diabetes, 37%, had heart failure, and 20% had prior stroke or transient ischemic attack. Many had high CHADS2 scores, indicating substantial risk for ischemic stroke and other adverse events. Anticoagulation control in the overall population was fair (mean TTR, 62.4%). Patients in the lowest tertile had TTR <55.6%; those in the highest tertile had TTR >73.4%.

Simulation of Improved Control and Impact on Adverse Events

In the status quo analysis, we simulated adverse events for all 67 077 patients according to their actual observed TTR (Table 5). During the 2-year study, 23.1% of patients (95% confidence interval [CI], 22.8 to 23.4) had any adverse event. Improving TTR by 2.5% averted 584 adverse events (95% CI, 302 to 883). These included 96 ischemic strokes (95% CI, −74 to 258), 136 major hemorrhages (95% CI, −24 to 289), and 352 deaths (95% CI, 126 to 584). Improving TTR by 5% averted 1114 adverse events (95% CI, 825 to 1389). These included 183 ischemic strokes (95% CI, 8 to 346), 269 major hemorrhages (95% CI, 118 to 416), and 662 deaths (95% CI, 435 to 869). Larger improvements in TTR were associated with even greater benefits.

Costs and Utilities/Sensitivity Analyses

Cost savings and utility gains from different levels of TTR improvement are shown in Table 5. Increasing TTR by 5% saved $15.9 million (95% CI, 3.2 to 28.4) and gained 863 QALYs (95% CI, 575 to 1154). Larger improvements in TTR had progressively greater benefits in terms of cost savings and utility gained. We also found that a QI program for oral anticoagulation is very likely to be cost-effective, although the cost of the program and its effectiveness in improving TTR cannot be known in advance. For example, if the QI program achieved a 5% improvement in TTR (a modest result), the program could cost as much as $59 million to implement and still achieve cost-effectiveness at the traditional $50 000/QALY threshold given our 2-year time horizon.

One-way sensitivity analyses are summarized in the online-only Data Supplement Appendix. Varying the cost

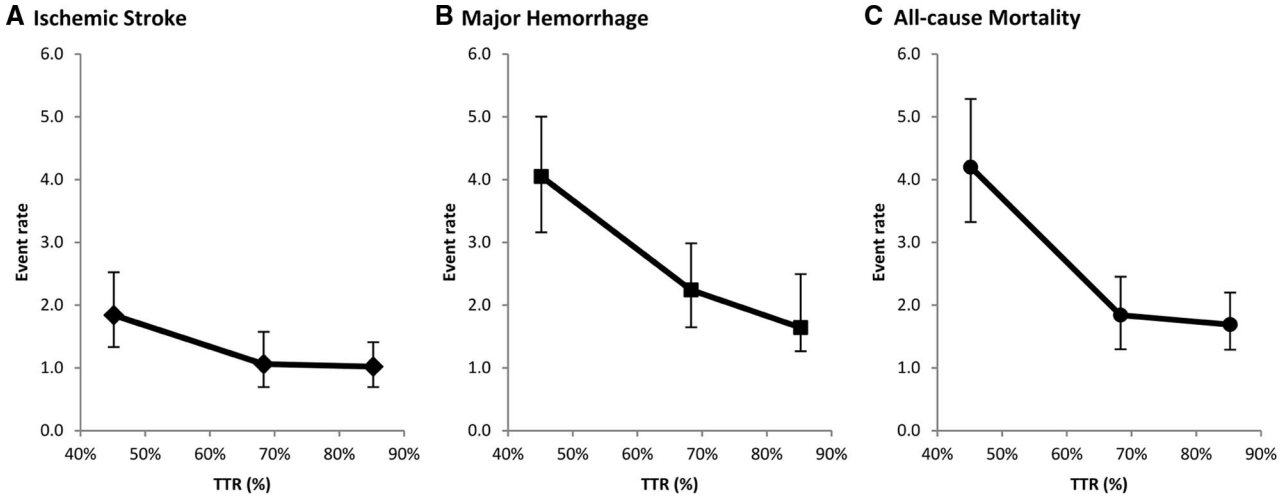


Figure 1. A, Interpolated relationships between TTR level and rates of ischemic stroke; **B**, major hemorrhage, and **C**, all-cause mortality. Rates are given in events per 100 patient-years.

Table 3. Cost and Utility Estimates (and Ranges) for Various Health Events and Health States Related to AF and Anticoagulation

Health Event or Health State	Estimate	Range
Costs		
Warfarin, 1 y	$210	$70–$400
Anticoagulation monitoring, 1 y; 17.4 visits/y	$520	$100–$1000
One-time cost of ischemic stroke		
Moderate-severe sequelae	$14 100	$7000–$30 000
Minor sequelae	$8700	$4000–$17 000
No sequelae	$7100	$4000–$14 000
Monthly (long-term) cost of ischemic stroke		
Moderate-severe sequelae	$5100	$2500–$10 000
Minor sequelae	$2300	$1000–$5000
One-time cost of ICH	$36 600	$17 000–$70 000
Monthly (long-term) cost of ICH	$5400	$2500–$10 000
One-time cost of non-ICH major hemorrhage	$4200	$2000–$7000
Utilities		
Healthy, using warfarin	0.987	0.953–1.0
Neurologic event with sequelae		
Moderate-severe sequelae (including ICH)	0.39	0.2–0.6
Minor sequelae	0.75	0.6–0.9
Major hemorrhage other than ICH	0.80	0.6–0.9

ICH indicates intracranial hemorrhage.

Costs are given in 2008 dollars. Utilities are given in QALYs and are reported per patient per year. One QALY is a full year of life in perfect health; a state of death is worth 0 QALYs, and a year of life with less than perfect health is valued as stated in the table. A major hemorrhage other than ICH causes the listed decrement in utility for only 1 month after the acute event, whereas strokes with sequelae cause the listed decrement in utility until the end of the study.

inputs did not meaningfully alter our results, with the possible exception of the monthly cost of an ischemic stroke with moderate-severe sequelae (example: increasing TTR by 5% saved between $12.2 to $22.7 million, depending on this one input). Varying the utility inputs did not meaningfully alter our results regarding expected utility gains. Varying the relationships between CHADS2 scores and events, or between TTR improvement and events, had the greatest impact on results. Gradients for ischemic stroke had the greatest impact on cost savings: For example, the savings from improving TTR by 5% varied from $8.3 to $35.3 million, depending on the relationship with CHADS2 and from $4.8 to $26.7 million, depending on the relationship with TTR. Gradients for death had the greatest impact on utility gains; for example, the gains from improving TTR by 5% varied from 532 to 1378 QALYs, depending on the relationship with CHADS2 and from 338 to 1230 QALYs, depending on the relationship with TTR. In the most extreme version of this sensitivity analysis, we assumed that changes in TTR had no impact on the risk of death. Predictably, this greatly reduced utility gains; for example, improving TTR by 5% gained 863 QALYs under the base case scenario and only 44 QALYs under this alternative scenario. Cost savings were largely unaffected by this maneuver, leaving the business case for quality improvement intact.

We also performed 2-way sensitivity analyses among the 3 variables with the largest impact on cost in the 1-way sensitivity analyses, including the relationship between CHADS2 score and stroke risk, the relationship between TTR and stroke risk, and the monthly cost of caring for a patient with moderate-severe stroke sequelae. In general, only extreme changes in these inputs had a sizeable impact on the overall cost savings. For example, the combined relationship between TTR, CHADS2, and stroke risk reduced estimated cost savings by >33% only if either improving TTR had no effect on stroke risk regardless of the relationship between CHADS2 and stroke risk or else if both improving TTR and higher CHADS2 scores had little effect on stroke risk.

Probabilistic sensitivity analyses confirmed that improving TTR is likely to result in benefit, although the exact size of the benefit might vary (Figure 2). For example, improving TTR by 5% was 76% likely to result in a savings of ≥$10 million and 97% likely to result in a utility gain of ≥500 QALYs. Of note, a very small improvement in TTR (2.5%) failed to save money approximately 6% of the time. Therefore, to ensure a high probability of being cost saving, a QI program would need to improve TTR by ≥3%.

Discussion

Improving the quality of oral anticoagulation for patients with AF can avert adverse events including ischemic stroke, major hemorrhage, and death.[11] We therefore simulated the effects of a program to improve anticoagulation control in the VA. In this population, even a modest improvement in TTR (5%)

Table 4. Baseline Sample Characteristics for 67 077 Patients Receiving Anticoagulation for AF in the VA

Variable	n (%)
Female sex	825 (1.2%)
Mean age, y (SD)	72.7 (9.6)
Race/ethnicity	
Non-Hispanic white	53 186 (79.3%)
Non-Hispanic black	3968 (5.9%)
Hispanic	1853 (2.8%)
Asian	226 (0.3%)
Native American	158 (0.2%)
Other/unknown	7686 (11.5%)
Comorbid conditions	
Coronary artery disease	36 052 (53.8%)
Diabetes	27 428 (40.9%)
Heart failure	24 974 (37.2%)
Hypertension	57 970 (86.4%)
Prior stroke or TIA	13 606 (20.3%)
CHADS2 risk score	
0	2256 (3.4%)
1	11 929 (17.8%)
2	21 604 (32.2%)
3	17 086 (25.5%)
4	8527 (12.7%)
5	4410 (6.6%)
6	1265 (1.9%)
Anticoagulation control	
Percent time in range (TTR), mean, SD	62.4% (21.5%)
Percent time in range (TTR), tertiles	<55.6%; 55.6%–73.4%, >73.4%

TIA indicates transient ischemic attack.

All patients had received anticoagulation for at least 6 months as of study inception.

would be expected to avert 1114 adverse events over 2 years, many of them fatal. We estimated that this change would result in a savings of $15.9 million and a gain of 863 QALYs over our 2-year study period. Even for a system as large as the VA, the cost of QI programs does not approach tens of millions of dollars. For example, the highly successful Translating Initiatives in Depression into Effective Solutions (TIDES) program has greatly improved management and outcomes for VA patients with depression.[42] The total cost of creating and implementing TIDES was recently estimated at $282 000.[43] Therefore, it seems likely that there is a compelling business case for our proposed QI initiative. Although our study is based on a VA population, its lesson would seem to apply to any integrated health care organization.

Anticoagulation therapy for AF is currently undergoing a major change, as dabigatran, the first serious competitor to warfarin, has recently received Food and Drug Administration approval. The RE-LY study showed that dabigatran is slightly superior to warfarin; for example, one would need to treat 357 patients with dabigatran (150 mg) to prevent 1 more ischemic stroke than warfarin would have prevented.[44] Although dabigatran is expected to improve patient convenience, it will be much more expensive than warfarin, even considering the cost of monitoring warfarin therapy.[45,46] It is likely that the small improvement in outcomes associated with dabigatran can be matched, or nearly matched, by a quality improvement program such as the one simulated here, but at a fraction of the cost. In other words, it may be easier to make a business case for QI in anticoagulation than for dabigatran. Our group plans to examine this issue.

We may have underestimated the benefits of improved TTR because of several of our assumptions. First, we only examined the impact of improved TTR for patients anticoagulated for AF because this is the condition for which we best understand the relationship between TTR and outcomes.[11] However, only about half of all patients receiving anticoagulation in the VA have AF,[12] and patients anticoagulated for other indications also achieve superior outcomes with improved control.[47,48] Any program to improve TTR would benefit these patients as well, though the magnitude of cost savings or cost effectiveness remains to be determined.

Second, given the available trial-based data on effectiveness,[18–35] we limited our study to a 2-year time horizon. However, preventing events such as ischemic stroke or major hemorrhage often provides cost and utility benefits for more than 2 years, another source of underestimated benefit. A third conservative feature of this study is that we censored patients after a first adverse event because this is how study outcomes are generally reported.[18–35] However, patients can have several adverse events, leading to additional costs. A fourth assumption is that we did not consider transient

Table 5. Simulated Adverse Events, Costs, and Utilities Over a 2-Year Period for the Entire Population of VA Patients With AF (n=67 077)

Scenario	Ischemic Strokes (95% CI)	Major Hemorrhages (95% CI)	Deaths (95% CI)	Total Events (95% CI)	Cost, $ Millions (95% CI)	Utility, QALYs (95% CI)
Status Quo	3872 (3749–3994)	3318 (3218–3429)	8285 (8109–8449)	15 475 (15 275–15 676)	368 (358–377)	121 685 (121 472–121 888)
Change in TTR	Estimated Adverse Events Averted, Costs Saved, or QALYs Added by Improving TTR					
+2.5%	96 (−74 to 258)	136 (−24 to 289)	352 (126–584)	584 (302, 883)	8.1 (−4.0–20.9)	458 (161–748)
+5%	183 (8–346)	269 (118–416)	663 (435–869)	1115 (825–1389)	15.9 (3.2–28.4)	863 (575–1154)
+7.5%	259 (91–421)	396 (251–548)	964 (722–1194)	1619 (1315–1912)	22.9 (10.9–35.2)	1255 (962–1528)
+10%	332 (171–499)	523 (377–685)	1233 (1000–1457)	2088 (1811–2364)	29.7 (16.7–41.7)	1606 (1338–1900)
+15%	454 (280–623)	755 (607–910)	1711 (1480–1931)	2920 (2630–3200)	41.3 (28.9–54.3)	2224 (1970–2224)
+20%	557 (394–720)	965 (826–1120)	2092 (1857–2307)	3614 (3320–3882)	51.5 (39.5–63.7)	2721 (2430–2983)

95% CIs are calculated by bootstrapping with 1000 iterations. Utilities are given in QALYs. Some of the estimates of benefit from the smallest TTR improvement that we modeled (2.5%) cannot be distinguished from zero at the 95% confidence level.

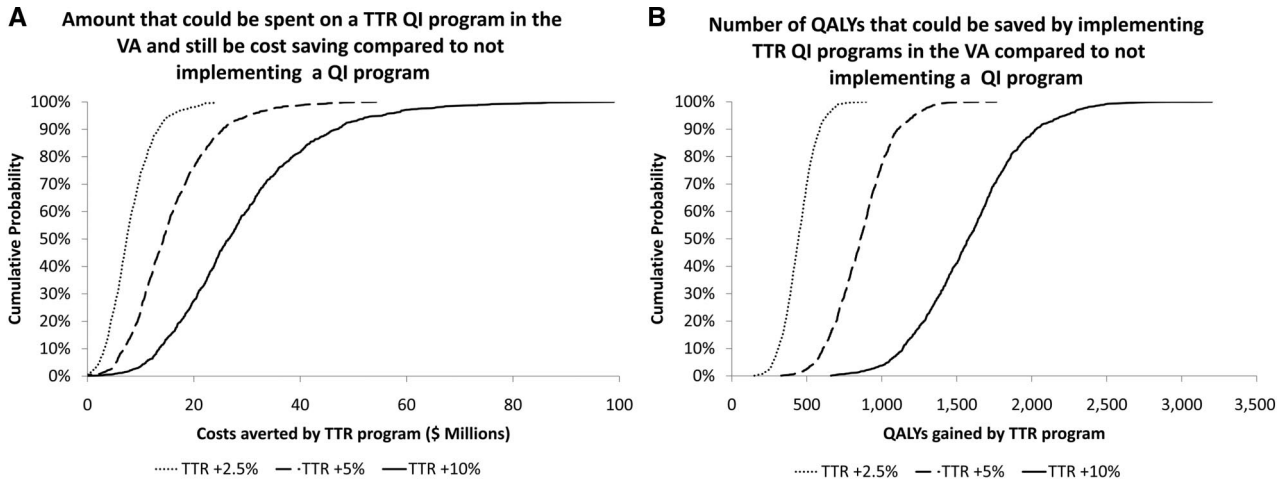


Figure 2. Distribution of outcomes from probabilistic sensitivity analyses, which varied all model inputs over their plausible ranges for 1000 iterations. These plots show the cumulative probability that (**A**) QI program for TTR would save a particular amount of money or (**B**) gain a particular number of QALYs.

ischemic attacks or minor hemorrhages because most trials do not report these events. Improved TTR should reduce the rates of these events as well. Finally, we did not consider the fact that improved TTR is likely not only to reduce the number of adverse events but also the severity of the events that are not prevented. We did not consider these effects because we lacked solid estimates for effect sizes, but their omission would also tend to underestimate the benefits of improved TTR.

This study has important strengths. First, as mentioned above, a number of underlying assumptions suggest that we have, if anything, underestimated the benefits of improved TTR. Second, we used a large data base of VA patients about which we have relatively complete data, especially regarding their CHADS2 risk for ischemic stroke. This allowed us to perform a detailed patient-level simulation to examine the impact of different degrees of TTR improvement in this population. Finally, much is known about the relationship between anticoagulation control and outcomes in AF, which is why we chose to focus on this condition in particular.

However, our study also has limitations. First, patient comorbidities were assessed using ICD-9 scores, which can be inaccurate. However, the CHADS2 scores that we ascribed to our population are well within the realm of what has been reported previously.[18–35] Second, we used CHADS2 not only as it was originally intended that is, as a risk score for ischemic stroke,[14] but also as a risk score for major hemorrhage and all-cause mortality; however, some literature supports using CHADS2 as a risk score for major hemorrhage.[15–17] When we used CHADS2 to risk-stratify for mortality, we used it as an all-purpose composite comorbidity score, similar to others that have been used in many prior studies.[49] This was necessary in large part because most of our source studies only reported these comorbidities.[18–35] Third, VA cost data are somewhat unique in that the VA does not usually bill for services. Therefore, we chose to use Medicare cost estimates for our study[37] to enhance generalizability to most settings. Fourth, our utility estimates are derived from a general population[37] rather than a VA population, and we did not include the inherent disutilities of comorbid conditions, old age, and ill health. This omission somewhat overstates benefits, although the extent of this is limited by our 2-year time horizon. Fifth, our estimates for event rates by CHADS2 score were derived from meta-analysis at the study level; however, in our data set we applied these coefficients at the individual patient level. To the extent that components of the CHADS2 score are correlated within individuals, our study-level analysis may in fact underestimate the risk of events with increasing CHADS2 scores at the individual level.

In conclusion, we modeled the possible benefits of a program to improve anticoagulation control for patients with AF in a large, integrated health care system. We found that such a program would be cost-saving even if it had only a minimal impact on control and even if it were considerably more expensive to implement than most QI programs have historically been. Our results suggest that the VA and other integrated healthcare systems should strongly consider implementing such a program. It is unusual to have an opportunity to save money while improving patient outcomes; this study describes such an opportunity.

Sources of Funding

This study was supported by a United States Department of Veterans Affairs, Health Services Research and Development Service, Career Development Award CDA-08-017 to Dr Rose. The sponsor had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, and approval of the manuscript.

Disclosures

Dr Hylek received honoraria from Bayer and Bristol Myers Squibb and serves on advisory boards for Boehringer-Ingelheim, Bristol Myers Squibb, Merck, and Sanofi Aventis.

References

1. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation: a major contributor to stroke in the elderly: the Framingham Study. *Arch Intern Med*. 1987; 147:1561–1564.
2. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke*. 1991;22:983–988.

3. Fuster V, Ryden LE, Cannom DS, Crijns HJ, Curtis AB, Ellenbogen KA, Halperin JL, Le Heuzey JY, Kay GN, Lowe JE, Olsson SB, Prystowsky EN, Tamargo JL, Wann S, Smith SC Jr, Jacobs AK, Adams CD, Anderson JL, Antman EM, Halperin JL, Hunt SA, Nishimura R, Ornato JP, Page RL, Riegel B, Priori SG, Blanc JJ, Budaj A, Camm AJ, Dean V, Deckers JW, Despres C, Dickstein K, Lekakis J, McGregor K, Metra M, Morais J, Osterspey A, Tamargo JL, Zamorano JL. ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation. *Circulation*. 2006;114:e257–e354.
4. Hart RG, Pearce LA, Aguillar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Ann Intern Med*. 2007;146:857–867.
5. Singer DE, Albers GW, Dalen JE, Fang MC, Go AS, Halperin JL, Lip GY, Manning WJ. Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th edition). *Chest.* 2008;133:546S–592S.
6. van Walraven C, Hart RG, Singer DE, Laupacis A, Connolly S, Petersen P, Koudstaal PJ, Chang Y, Hellemons B. Oral anticoagulants vs aspirin in nonvalvular atrial fibrillation: an individual patient meta-analysis. *JAMA*. 2002;288:2441–2448.
7. van Walraven C, Jennings A, Oake N, Fergusson D, Forster AJ. Effect of study setting on anticoagulation control: a systematic review and metaregression. *Chest*. 2006;129:1155–1166.
8. Rose AJ, Berlowitz DR, Frayne SM, Hylek EM. Measuring quality of oral anticoagulation care: extending quality measurement to a new field. *Jt Comm J Qual Patient Saf*. 2009;35:146–155.
9. Rose AJ, Hylek EM, Ozonoff A, Ash AS, Reisman JI, Berlowitz DR. Risk-adjusted percent time in therapeutic range as a quality indicator for outpatient oral anticoagulation: results of the veterans affairs study to improve anticoagulation (varia). *Circ Cardiovasc Qual Outcomes*. 2011; 4:22–29.
10. Connolly SJ, Pogue J, Eikelboom J, Flaker G, Commerford P, Franzosi MG, Healey JS, Yusuf S. Benefit of oral anticoagulant over antiplatelet therapy in atrial fibrillation depends on the quality of international normalized ratio control achieved by centers and countries as measured by time in therapeutic range. *Circulation*. 2008;118:2029–2037.
11. White HD, Gruber M, Feyzi J, Kaatz S, Tse HF, Husted S, Albers GW. Comparison of outcomes among patients randomized to warfarin therapy according to anticoagulant control: results from SPORTIF III and V. *Arch Intern Med*. 2007;167:239–245.
12. Rose AJ, Hylek EM, Ozonoff A, Ash AS, Reisman JI, Berlowitz DR. Patient characteristics associated with oral anticoagulation control: results of the Veterans Affairs Study to Improve Anticoagulation (VARIA). *J Thromb Haemost*. 2010;8:2182–2191.
13. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briet E. A method to determine the optimal intensity of oral anticoagulant therapy. *Thromb Haemost*. 1993;69:236–239.
14. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA*. 2001; 285:2864–2870.
15. Healey JS, Hart RG, Pogue J, Pfeffer MA, Hohnloser SH, De Caterina R, Flaker G, Yusuf S, Connolly SJ. Risks and benefits of oral anticoagulation compared with clopidogrel plus aspirin in patients with atrial fibrillation according to stroke risk: the Atrial Fibrillation Clopidogrel Trial With Irbesartan for Prevention of Vascular Events (ACTIVE-W). *Stroke*. 2008;39:1482–1486.
16. Hylek EM, Evans-Molina C, Shea C, Henault LE, Regan S. Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. *Circulation*. 2007;115:2689–2696.
17. Poli D, Antonucci E, Marcucci R, Fatini C, Alterini B, Mannini L, Falciani M, Abbate R, Gensini GF, Prisco D. Risk of bleeding in very old atrial fibrillation patients on warfarin: relationship with ageing and chads2 score. *Thromb Res*. 2007;121:347–352.
18. The effect of low-dose warfarin on the risk of stroke in patients with nonrheumatic atrial fibrillation: the Boston Area Anticoagulation Trial for Atrial Fibrillation Investigators. *N Engl J Med.* 1990;323:1505–1511.
19. Stroke prevention in atrial fibrillation study: final results. *Circulation.* 1991;84:527–539.
20. Secondary prevention in non-rheumatic atrial fibrillation after transient ischaemic attack or minor stroke: EAFT (European Atrial Fibrillation Trial) study group. *Lancet.* 1993;342:1255–1262.
21. Warfarin versus aspirin for prevention of thromboembolism in atrial fibrillation: Stroke Prevention in Atrial Fibrillation II Study. *Lancet.* 1994;343:687–691.
22. Adjusted-dose warfarin versus low-intensity, fixed-dose warfarin plus aspirin for high-risk patients with atrial fibrillation: Stroke Prevention in Atrial Fibrillation III randomised clinical trial. *Lancet.* 1996;348: 633–638.
23. Albers GW, Diener HC, Frison L, Grind M, Nevinson M, Partridge S, Halperin JL, Horrow J, Olsson SB, Petersen P, Vahanian A. Ximelagatran vs warfarin for stroke prevention in patients with nonvalvular atrial fibrillation: a randomized trial. *JAMA*. 2005;293:690–698.
24. Connolly S, Pogue J, Hart R, Pfeffer M, Hohnloser S, Chrolavicius S, Yusuf S. Clopidogrel plus aspirin versus oral anticoagulation for atrial fibrillation in the Atrial Fibrillation Clopidogrel Trial With Irbesartan for Prevention of Vascular Events (ACTIVE W): a randomised controlled trial. *Lancet*. 2006;367:1903–1912.
25. Connolly SJ, Laupacis A, Gent M, Roberts RS, Cairns JA, Joyner C. Canadian atrial fibrillation anticoagulation (CAFA) study. *J Am Coll Cardiol*. 1991;18:349–355.
26. Ezekowitz MD, Bridgers SL, James KE, Carliner NH, Colling CL, Gornick CC, Krause-Steinrauf H, Kurtzke JF, Nazarian SM, Radford MJ, Rickles FR, Shabetai R, Deykin D. Warfarin in the prevention of stroke associated with nonrheumatic atrial fibrillation: Veterans Affairs Stroke Prevention in Nonrheumatic Atrial Fibrillation Investigators. *N Engl J Med*. 1992;327:1406–1412.
27. Gullov AL, Koefoed BG, Petersen P. Bleeding during warfarin and aspirin therapy in patients with atrial fibrillation: the AFASAK 2 study: Atrial Fibrillation Aspirin and Anticoagulation. *Arch Intern Med*. 1999; 159:1322–1328.
28. Hellemons BS, Langenberg M, Lodder J, Vermeer F, Schouten HJ, Lemmens T, van Ree JW, Knottnerus JA. Primary prevention of arterial thromboembolism in non-rheumatic atrial fibrillation in primary care: randomised controlled trial comparing two intensities of coumarin with aspirin. *BMJ*. 1999;319:958–964.
29. Morocutti C, Amabile G, Fattapposta F, Nicolosi A, Matteoli S, Trappolini M, Cataldo G, Milanesi G, Lavezzari M, Pamparana F, Coccheri S. Indobufen versus warfarin in the secondary prevention of major vascular events in nonrheumatic atrial fibrillation: SIFA (Studio Italiano Fibrillazione Atriale) Investigators. *Stroke*. 1997;28:1015–1021.
30. Olsson SB. Stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation (SPORTIF III): randomised controlled trial. *Lancet*. 2003;362: 1691–1698.
31. Pengo V, Zasso A, Barbero F, Banzato A, Nante G, Parissenti L, John N, Noventa F, Dalla Volta S. Effectiveness of fixed minidose warfarin in the prevention of thromboembolism and vascular death in nonrheumatic atrial fibrillation. *Am J Cardiol*. 1998;82:433–437.
32. Go AS, Hylek EM, Chang Y, Phillips KA, Henault LE, Capra AM, Jensvold NG, Selby JV, Singer DE. Anticoagulation therapy for stroke prevention in atrial fibrillation: how well do randomized trials translate into clinical practice? *JAMA*. 2003;290:2685–2692.
33. Mant J, Hobbs FD, Fletcher K, Roalfe A, Fitzmaurice D, Lip GY, Murray E. Warfarin versus aspirin for stroke prevention in an elderly community population with atrial fibrillation (the Birmingham Atrial Fibrillation Treatment of the Aged Study, BAFTA): a randomised controlled trial. *Lancet*. 2007;370:493–503.
34. Petersen P, Boysen G, Godtfredsen J, Andersen ED, Andersen B. Placebo-controlled, randomised trial of warfarin and aspirin for prevention of thromboembolic complications in chronic atrial fibrillation: the Copenhagen AFASAK study. *Lancet*. 1989;1:175–179.
35. Rose AJ, Ozonoff A, Henault LE, Hylek EM. Warfarin for atrial fibrillation in community-based practise. *J Thromb Haemost*. 2008;6: 1647–1654.
36. Amouyel P, Mismetti P, Langkilde LK, Jasso-Mosqueda G, Nelander K, Lamarque H. INR variability in atrial fibrillation: a risk model for cerebrovascular events. *Eur J Intern Med*. 2009;20:63–69.
37. O'Brien CL, Gage BF. Costs and effectiveness of ximelagatran for stroke prophylaxis in chronic atrial fibrillation. *JAMA*. 2005;293:699–706.
38. Consumer price index data bases. http://www.Bls.Gov/cpi/. Accessed march 21, 2011.
39. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-Effectiveness in Health and Medicine.* New York: Oxford University Press; 1996.
40. Jones M, McEwan P, Morgan CL, Peters JR, Goodfellow J, Currie CJ. Evaluation of the pattern of treatment, level of anticoagulation control, and outcome of treatment with warfarin in patients with non-valvar atrial fibrillation: a record linkage study in a large British population. *Heart*. 2005;91:472–477.

41. Morgan CL, McEwan P, Tukiendorf A, Robinson PA, Clemens A, Plumb JM. Warfarin treatment in patients with atrial fibrillation: observing outcomes associated with varying levels of INR control. *Thromb Res*. 2009;124:37–41.
42. Rubenstein LV, Chaney EF, Ober S, Felker B, Sherman SE, Lanto A, Vivell S. Using evidence-based quality improvement methods for translating depression collaborative care research into practice. *Fam Syst Health*. 2010;28:91–113.
43. Liu CF, Rubenstein LV, Kirchner JE, Fortney JC, Perkins MW, Ober SK, Pyne JM, Chaney EF. Organizational cost of quality improvement for depression care. *Health Serv Res*. 2009;44:225–244.
44. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, Pogue J, Reilly PA, Themeles E, Varrone J, Wang S, Alings M, Xavier D, Zhu J, Diaz R, Lewis BS, Darius H, Diener HC, Joyner CD, Wallentin L. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2009;361:1139–1151.
45. Freeman JV, Zhu RP, Owens DK, Garber AM, Hutton DW, Go AS, Wang PJ, Turakhia MP. Cost-effectiveness of dabigatran compared with warfarin for stroke prevention in atrial fibrillation. *Ann Intern Med.* 2010;154:1–11.
46. Patel R, Medow M. Cost-effectiveness based pricing of dabigatran for the primary prophylaxis of stroke in elderly patients with atrial fibrillation (abstract). Presented at the 2010 annual meeting of the Society of General Internal Medicine, Minneapolis, MN. April 28, 2010.
47. van Leeuwen Y, Rosendaal FR, Cannegieter SC. Prediction of hemorrhagic and thrombotic events in patients with mechanical heart valve prostheses treated with oral anticoagulants. *J Thromb Haemost*. 2008;6:451–456.
48. Veeger NJ, Piersma-Wichers M, Tijssen JG, Hillege HL, van der Meer J. Individual time within target range in patients treated with vitamin k antagonists: main determinant of quality of anticoagulation and predictor of clinical outcome: a retrospective study of 2300 consecutive patients with venous thromboembolism. *Br J Haematol*. 2005;128:513–519.
49. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-cm administrative data bases. *J Clin Epidemiol*. 1992;45:613–619.

SUPPLEMENTAL MATERIAL
Online Appendix: Results of deterministic sensitivity analyses

Table A1: One-way deterministic sensitivity analyses for varying cost inputs. Cost savings, compared to the status quo, are given in millions of dollars over a two-year period.

	Lowest	Low	Middle	High	Highest
Cost of Warfarin					
Status Quo	--	--	--	--	--
TTR 5	18.4	16.2	15.9	16.1	16.6
TTR 10	30.9	30.1	29.7	30.8	31.3
TTR 15	41.9	41.0	41.3	40.4	40.9
Cost of OAT Monitoring					
Status Quo	--	--	--	--	--
TTR 5	17.4	18.5	15.9	15.3	16.9
TTR 10	29.2	31.3	29.7	28.2	30.1
TTR 15	42.3	43.1	41.3	39.3	40.6
One-Time Cost of Moderate-Severe Stroke					
Status Quo	--	--	--	--	--
TTR 5	15.0	15.3	15.9	16.5	15.2
TTR 10	29.5	29.2	29.7	31.6	28.7
TTR 15	41.5	41.4	41.3	43.0	43.7
One-Time Cost of Mild Stroke					
Status Quo	--	--	--	--	--
TTR 5	15.0	15.3	15.9	16.5	15.2
TTR 10	29.5	29.2	29.7	31.6	28.7
TTR 15	41.5	41.4	41.3	43.0	43.7
One-Time Cost of Stroke Without Sequellae					
Status Quo	--	--	--	--	--
TTR 5	15.0	15.3	15.9	16.5	15.2
TTR 10	29.5	29.2	29.7	31.6	28.7
TTR 15	41.5	41.4	41.3	43.0	43.7
Monthly Cost of Moderate-Severe Stroke					
Status Quo	--	--	--	--	--
TTR 5	12.2	16.0	15.9	19.9	22.7
TTR 10	24.2	26.0	29.7	36.2	38.0
TTR 15	32.5	38.6	41.3	52.0	54.2
Monthly Cost of Mild Stroke					
Status Quo	--	--	--	--	--
TTR 5	14.3	15.5	15.9	17.4	17.6

TTR 10	26.8	29.0	29.7	33.4	35.3
TTR 15	38.0	38.0	41.3	44.4	49.2
One-Time Cost of ICH					
Status Quo	--	--	--	--	--
TTR 5	14.4	15.1	15.9	16.8	16.6
TTR 10	27.7	27.6	29.7	30.7	32.4
TTR 15	37.3	39.9	41.3	44.4	46.5
Monthly Cost of ICH					
Status Quo	--	--	--	--	--
TTR 5	14.4	14.9	15.9	15.3	16.6
TTR 10	24.9	27.9	29.7	29.5	32.4
TTR 15	36.8	36.4	41.3	43.3	46.5
One-Time Cost of Non-ICH Major Hemorrhage					
Status Quo	--	--	--	--	--
TTR 5	16.2	17.4	15.9	16.9	16.3
TTR 10	30.0	29.6	29.7	29.7	29.5
TTR 15	39.9	42.2	41.3	41.9	42.9

TTR 5: Percent time in range improved by 5%
TTR 10: Percent time in range improved by 10%
TTR 15: Percent time in range improved by 15%
Trends are not always monotonic across categories because in some cases, the variability inherent in the random number simulation exceeds the differences between categories.

Table A2: One-way deterministic sensitivity analyses for varying utility inputs. Utility gains, compared to the status quo, are given in QALYs over a two-year period.

	Lowest	Low	Middle	High	Highest
Healthy Utility					
Status Quo	--	--	--	--	--
TTR 5	808	833	863	880	861
TTR 10	1521	1575	1606	1599	1604
TTR 15	2140	2162	2224	2220	2228
Utility of Moderate-Severe Stroke (Including ICH)					
Status Quo	--	--	--	--	--
TTR 5	890	846	863	838	855
TTR 10	1619	1641	1606	1572	1532
TTR 15	2297	2270	2224	2152	2134
Utility of Mild Stroke					
Status Quo	--	--	--	--	--
TTR 5	874	895	863	902	851
TTR 10	1627	1659	1606	1591	1588
TTR 15	2271	2254	2224	2222	2182
Utility of Non-ICH Bleed (for 1 Month)					
Status Quo	--	--	--	--	--
TTR 5	881	846	863	893	858
TTR 10	1614	1573	1606	1630	1534
TTR 15	2220	2219	2224	2241	2194

TTR 5: Percent time in range improved by 5%
TTR 10: Percent time in range improved by 10%
TTR 15: Percent time in range improved by 15%
Trends are not always monotonic across categories because in some cases, the variability inherent in the random number simulation exceeds the differences between categories.

Table A3: One-way deterministic sensitivity analyses for varying the relationships between $CHADS^2$ score and the risk of adverse events, and for varying the relationship between improved anticoagulation control and the risk of adverse events. Cost savings, compared to the status quo, are given in millions of dollars over a two-year period. Utility gains, compared to the status quo, are given in QALYs over a two-year period.

	Shallowest	Shallow	Middle	Steep	Steepest
$CHADS^2$ Gradient for Stroke					
Status Quo	--	--	--	--	--
TTR 5 - COST	8.3	11.4	15.9	22.8	35.3
TTR 10 - COST	17.1	21.2	29.7	45.8	61.9
TTR 15 - COST	24.9	30.3	41.3	62.8	88.8
Status Quo	--	--	--	--	--
TTR 5 - UTILITY	854	871	863	847	866
TTR 10 - UTILITY	1586	1619	1606	1566	1590
TTR 15 - UTILITY	2195	2214	2224	2176	2207
$CHADS^2$ Gradient for Hemorrhage					
Status Quo	--	--	--	--	--
TTR 5 - COST	13.1	15.3	15.9	17.9	19.0
TTR 10 - COST	26.6	26.2	29.7	34.1	39.1
TTR 15 - COST	33.5	38.2	41.3	50.1	56.8
Status Quo	--	--	--	--	--
TTR 5 - UTILITY	866	910	863	879	776
TTR 10 - UTILITY	1624	1609	1606	1532	1493
TTR 15 - UTILITY	2254	2281	2224	2192	2078
$CHADS^2$ Gradient for Death					
Status Quo	--	--	--	--	--
TTR 5 - UTILITY	532	650	863	1121	1378
TTR 10 - UTILITY	992	1218	1606	2101	2568
TTR 15 - UTILITY	1355	1697	2224	2848	3554
TTR Gradient for Stroke					
Status Quo	--	--	--	--	--
TTR 5 - COST	4.8	10.4	15.9	21.5	26.7
TTR 10 - COST	8.8	18.4	29.7	42.8	53.9
TTR 15 - COST	12.8	29.0	41.3	58.0	72.1
Status Quo	--	--	--	--	--
TTR 5 - UTILITY	844	852	863	915	926
TTR 10 - UTILITY	1507	1524	1606	1707	1726
TTR 15 - UTILITY	2100	2148	2224	2340	2347
TTR Gradient for Hemorrhage					
Status Quo	--	--	--	--	--

TTR 5 - COST	12.1	15.0	15.9	16.4	19.0
TTR 10 - COST	22.7	26.6	29.7	31.1	33.5
TTR 15 - COST	33.3	36.0	41.3	44.6	49.5
Status Quo	--	--	--	--	--
TTR 5 - UTILITY	845	862	863	902	855
TTR 10 - UTILITY	1615	1606	1606	1608	1577
TTR 15 - UTILITY	2203	2226	2224	2268	2174
TTR Gradient for Death					
Status Quo	--	--	--	--	--
TTR 5 - UTILITY	338	682	863	1032	1230
TTR 10 - UTILITY	622	1216	1606	1932	2262
TTR 15 - UTILITY	904	1682	2224	2657	3076

TTR 5: Percent time in range improved by 5%
TTR 10: Percent time in range improved by 10%
TTR 15: Percent time in range improved by 15%
$CHADS^2$: A six-point risk score for adverse events
Sensitivity analyses for TTR gradients are for the segment joining poor to moderate TTR; the segment between moderate and good TTR is less important (because it is very shallow) and is therefore not considered here. Trends are not always monotonic across categories because in some cases, the variability inherent in the random number simulation exceeds the differences between categories.

Figure A1: Tornado plot of cost inputs. Costs shown are the average cost per patient, in dollars.

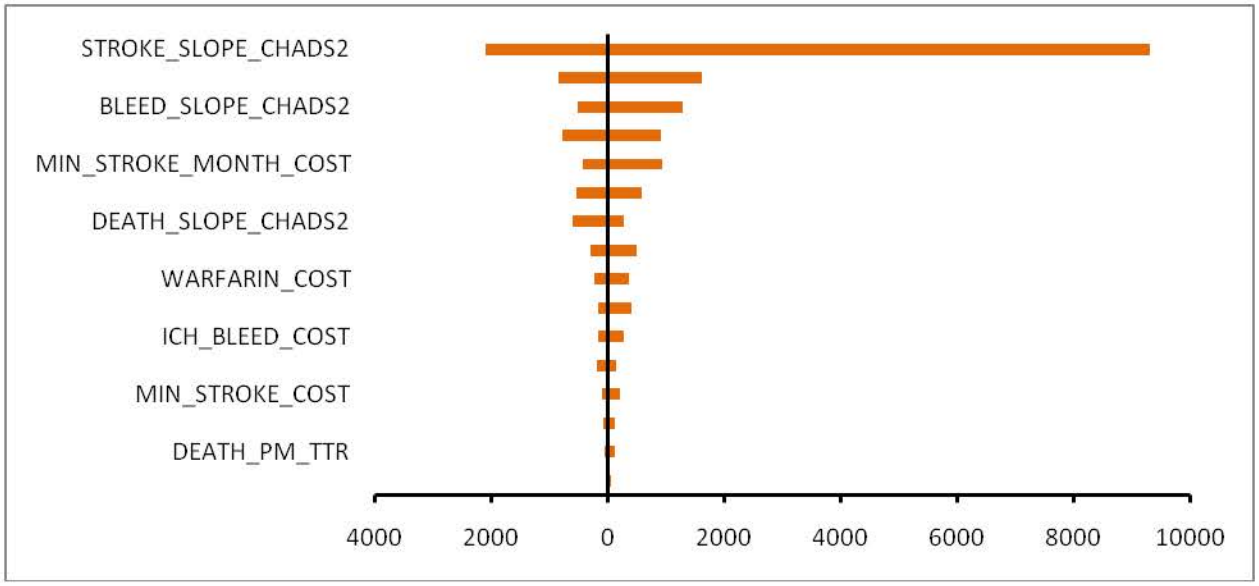


Figure A2: Tornado plot of utility inputs. Utilities shown in QALYs per patient.

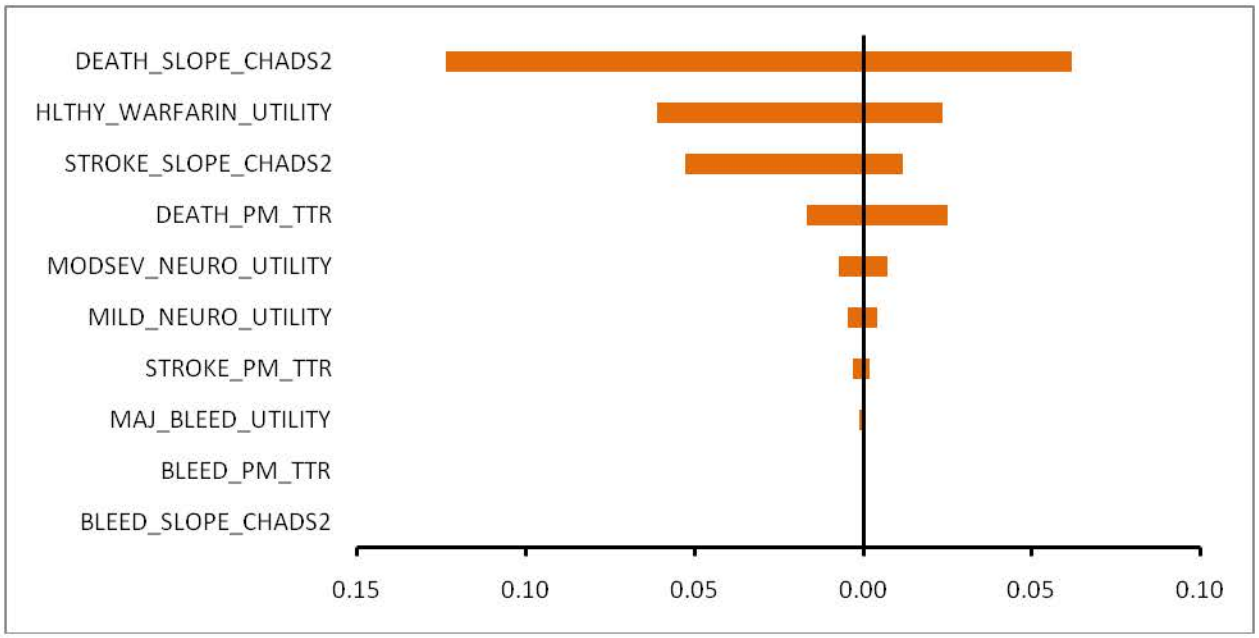


Table A4: Extreme sensitivity analysis. All costs or utilities are set to their lowest level, their highest level, a low-middle level, and a high-middle level. Cost savings, compared to the status quo, are given in millions of dollars over a two-year period. Utility gains, compared to the status quo, are given in QALYs over a two-year period.

	Low-Middle	Base Case	High-Middle
Cost Savings			
Status Quo	--	--	--
TTR 5	10.7	15.9	24.5
TTR 10	22.9	29.7	45.8
TTR 15	30.3	41.3	62.7
Utility Gained			
Status Quo	--	--	--
TTR 5	907	863	834
TTR 10	1624	1606	1587
TTR 15	2223	2224	2183

TTR 5: Percent time in range improved by 5%
TTR 10: Percent time in range improved by 10%
TTR 15: Percent time in range improved by 15%
Trends are not always monotonic across categories because in some cases, the variability inherent in the random number simulation exceeds the differences between categories.

Table A5: Extreme sensitivity analysis. Assume that improved anticoagulation control only reduces the rates of non-fatal events; death rates do not change at all. Cost savings, compared to the status quo, are given in millions of dollars over a two-year period. Utility gains, compared to the status quo, are given in QALYs over a two-year period.

Cost Savings

	Base Case	Sensitivity
Status Quo	--	--
TTR 5	15.9	18.8
TTR 10	29.7	35.7
TTR 15	41.3	48.2

Utility Gains

	Base Case	Sensitivity
Status Quo	--	--
TTR 5	863	44
TTR 10	1606	103
TTR 15	2224	188

TTR 5: Percent time in range improved by 5%
TTR 10: Percent time in range improved by 10%
TTR 15: Percent time in range improved by 15%

Effects of Daily Adherence to Antihypertensive Medication on Blood Pressure Control

Adam J. Rose, MD, MSc;[1,2] Mark E. Glickman, PhD;[1,3] Meredith M. D'Amore, MPH;[2] Michelle B. Orner, MPH;[1] Dan Berlowitz, MD, MPH;[1,2,3] Nancy R. Kressin, PhD[1,2,4]

From the Center for Health Quality, Outcomes, and Economics Research, Bedford VA Medical Center, Bedford;[1] the Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston;[2] the Department of Health Policy and Management, Boston University School of Public Health, Boston;[3] and the VA Boston Health Care System, Boston,[4] MA

Clinicians are often uncertain about how to manage elevated blood pressure (BP) when a patient reports that he/she has recently missed several doses of antihypertensive medications. While we know that better adherence can improve BP during several months, the magnitude of this relationship in the short term is poorly understood. The authors examined this issue using a group of patients who monitored adherence using a Medication Events Monitoring System (MEMS) cap and had BP measurements in the course of routine clinical practice. BP readings were compared following 7 days of excellent adherence (100%) or poor adherence (<60%), omitting BP values following intermediate adherence. Using several different methods, BP following 7 days of excellent adherence was between 12/7 mm Hg and 15/8 mm Hg lower than after 7 days of poor adherence. Clinicians can use this effect size to calibrate their impressions of what the BP might have been with improved adherence. *J Clin Hypertens (Greenwich).* 2011;13:416–421. ©2011 Wiley Periodicals, Inc.

Treatment adherence is an important determinant of blood pressure (BP) control.[1,2] In usual clinical practice, treatment adherence can vary over time. While many studies have categorized patients as "adherent" and "nonadherent,"[2–5] most patients have periods of better and worse adherence that do not fit into such dichotomized categories.[6–10] Clinicians are commonly informed by a patient that his or her adherence has been less than perfect during the past week because the patient ran out of medication, forgot, or was confused, or for other reasons. It is unclear how clinicians should interpret BP values obtained after the patient has reported a period of relatively poor adherence because there are few available estimates of the expected impact on BP. It would therefore be useful to estimate the impact of poor adherence on BP during a short period.

There are at least 3 kinds of previous studies that might shed some light on the impact of a brief period of poor adherence on BP, but each has shortcomings. Most medications, particularly at the time they are being considered for approval, are evaluated compared with placebo. This provides some information about the effect of a single medication under controlled circumstances. However, the applicability to clinical practice is limited because real-life regimens usually contain more than 1 drug, and real-life nonadherence is usually partial rather than complete.[6–10] There are a few studies in which patients are instructed to stop their antihypertensive medications abruptly in order to compare the rebound effects of different medication classes during 7 days.[11] While such a study can provide valuable physiologic data, it also simplifies regimens down to only 1 drug and does not mirror real-world patterns of nonadherence. Finally, there are large retrospective database studies in which adherence is usually characterized using pharmacy fill data[3–5] or patient self-report.[1,2,12] Such studies have often shown that so-called nonadherent patients have worse BP control than adherent patients. Shortcomings of such designs include the unspecified time relationship between adherence behavior and BP measurements, as well as the oversimplification inherent in dividing patients using a binary adherence measure (adherent vs nonadherent).[10,13]

Recently, we have had another tool for measuring adherence: the Medication Event Monitoring System [MEMS] caps (Aardex Group, Ltd, Sion, Switzerland). A MEMS cap records each bottle opening, allowing clinicians and researchers access to extremely detailed data regarding persistence with therapy and timeliness of dosing.[6–10,14–17] The availability of detailed adherence data from MEMS caps provides an opportunity to better characterize the effect on BP of a brief period of poor adherence in a real-world setting. We therefore examined data from a study on hypertension in which patients used MEMS caps to monitor adherence. We sought to characterize the precise relationship between a 7-day period of poor adherence to antihypertensive therapy and the resultant change in BP.

METHODS

Data

The data for our analyses were obtained from the pre-intervention period of a randomized trial that

Address for correspondence: Mark E. Glickman, PhD, Center for Health Quality, Outcomes and Economics Research, 200 Springs Road (152), Bedford, MA 01730
E-mail: mg@bu.edu

Manuscript received: July 28, 2010; **Revised:** November 22, 2010; **Accepted:** November 23, 2010
DOI: 10.1111/j.1751-7176.2011.00427.x

examined the effects of a provider-patient communication skill-building intervention on adherence to antihypertensive medication therapy and BP control (clinicaltrials.gov identifier: NCT00201149). Patients were enrolled from 7 outpatient primary care clinics at Boston Medical Center, an inner-city safety-net hospital affiliated with the Boston University School of Medicine. The study was approved by the institutional review board of Boston University Medical Center.

Patients were recruited from August 2004 through June 2006 if they were of white or black race, aged at least 21 years, and had an outpatient diagnosis of hypertension on at least 3 separate occasions prior to enrollment. Based on initial screening, patients were ineligible for the study if they already used a medication dispenser (as this might invalidate adherence data collection), were cognitively impaired, were of an ethnicity/race other than white or black, were unable to speak English, were not prescribed antihypertensive medication, were already participating in another hypertension study, or refused to participate.

Among 869 patients enrolled in the study who received dispensers with MEMS caps to monitor adherence to antihypertensive medication, 689 returned them. Our current study focuses on medication-taking behavior during the 90 days after the first opening of the MEMS cap. Rather than giving a patient multiple MEMS caps for all agents in their hypertension regimen, we gave one MEMS cap to each patient to correspond with one of their antihypertensive medications, asking them to use the MEMS cap for the most frequently taken medication. We characterized the medication-taking behavior of each patient using only this one medication, the ''index medication.'' A similar strategy has been pursued in prior studies using MEMS caps to characterize adherence to a multidrug regimen, which have found that adherence to an index medication generally correlates with adherence to the entire regimen.[10,16,17]

We imposed additional restrictions to increase the homogeneity of the analytic sample. During the first 90 days after issuance of the MEMS cap, a patient needed at least 2 clinic visits with BP readings to be part of our final sample. Ensuring multiple BP readings per patient reduced the potential for confounding effects of adherence and characteristics specific to a patient. This reduced the sample to 249 patients. We then excluded 35 patients taking regimens of $\geq$2 doses per day, one patient whose index medication changed during the first 90 days, and 3 others whose dose frequency for the index medication changed from twice to once daily during the first 90 days. This resulted in 210 patients who were taking regimens of 1 dose per day for the first 90 days of the study. Furthermore, if a patient opened his/her MEMS cap more than twice per day during $\geq$10% of the monitored period, then the patient was excluded because of suspicion that the patient did not understand the MEMS cap and was not using it correctly. This resulted in a final study sample of 200 patients.

Independent Variable: Adherence to Therapy

We characterized adherence to antihypertensive therapy using MEMS caps. These devices use a microchip to record all bottle openings. Good adherence as measured by MEMS caps has been linked to improvements in numerous clinical outcomes,[7,8,14,15] including hypertension control.[16,17] In the current study, clinicians were not given feedback about their patients' adherence as measured by MEMS caps.

The MEMS cap data for this sample were cleaned in the following manner. For nonmonitored periods (eg, hospitalizations), the number of MEMS cap openings were treated as missing. A patient was considered adherent on days in which the MEMS cap was recorded to have been opened exactly once or twice and was considered nonadherent if the MEMS cap was not opened. On days where the MEMS cap was opened more than twice, the number of openings was considered missing data due to the extra uncertainty in the reason for the multiple openings.

Dependent Variable: Clinical BP Measurements

BP was taken for each patient at irregular intervals, as part of routine clinical care. BPs could be taken using manual or electronic devices by clinical staff including physicians, nurses, and medical assistants and were recorded in the electronic medical record. If multiple readings were taken on a single day, the values were averaged for our study. We separately examined systolic BP (SBP) and diastolic BP (DBP) as outcomes.

Control Variables

We recorded sex, self-reported race (white vs black), and age at study inception. Using both *International Classification of Diseases, Ninth Revision* codes and problem lists from the electronic medical record, we noted whether the patients had the following comorbid conditions, all of which could impact BP, the use of antihypertensive medications, or the perceived urgency of controlling hypertension: cerebrovascular disease, congestive heart failure, chronic kidney disease, coronary artery disease, diabetes mellitus, hyperlipidemia, obesity (body mass index $>$30 kg/m^2), and peripheral vascular disease.

Patient Adherence Example

An example of the relationship between MEMS cap openings and BP measurements can be seen for one of the patients in our study (Figure). The horizontal axis counts the number of days since the patient entered the study, and the vertical axis counts the number of MEMS cap openings on each day displayed as black dots. For this patient, 3 clinic visits occurred in which BP readings were taken, with the days represented by the vertical lines and the SBP and DBP indicated by diamonds on the lines. The 7 days preceding the BP

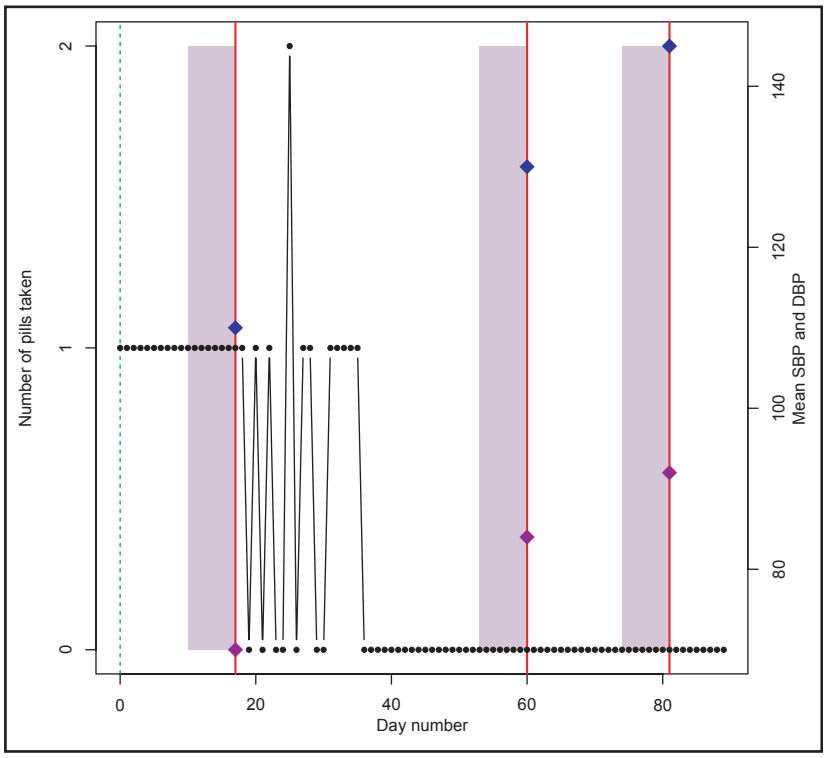


FIGURE. Example patient profile. The number of pills taken on each day of the study for a selected patient. The 3 vertical lines indicate clinic visit days on which blood pressure readings were taken, and the diamonds on each line denote the systolic blood pressure (SBP) and diastolic blood pressure (DBP) at the visit (labeled on the right). The shaded rectangular areas mark off 7 days prior to each clinic visit.

visits are shaded. In the 7-day period preceding the first visit, the patient had excellent adherence (1 pill per day), and the BP reading was 110/70 mm Hg. For the subsequent two visits, the adherence was at 0% for the 7-day period preceding the BP readings, and the corresponding BP measurements were 130/84 mm Hg on the first visit and 145/92 mm/Hg on the second visit. This example provides an informal basis for the statistical analyses we apply to these data.

Statistical Analysis

Two distinct approaches were taken to assess the effects of adherence on BP. The first approach treated each BP reading as a separate outcome and used adherence during the preceding 7 days to predict BP. For these analyses, we used the full analytic sample of 200 patients. The second approach restricted the sample to patients who had $\geq$2 BP readings, one of which was preceded by excellent adherence and one of which was preceded by poor adherence. With the first approach, we analyzed a larger sample and adjusted for between-patient differences through the control and health factors. In the second approach, every patient served as his/her own control and it was not necessary to control for patient-specific covariates. In both cases, recent adherence prior to a BP reading was determined based on the percentage of days in which the patient opened the MEMS cap. Days in which MEMS cap openings were missing were not counted in the adherence calculation. Adherence was considered poor if adherence was <60% and excellent if adherence was 100%. BP readings preceded by a 7-day period with intermediate adherence (between 60% and 100%) were removed from this analysis to provide a more precise estimate of the two ends of the scale, which resulted in a sample size decrease to 178. All models were separately fit to predict SBP and DBP.

In our first analytic approach, adherence was computed for the 7 days prior to each BP reading. BP readings were excluded if they occurred within the first 7 days of a patient's entry to the study. Random effects least-squares regressions were fit to the resulting data. BP (SBP and DBP) were regressed on an excellent/poor adherence indicator, along with sex, race, age (in years, categorized into 0–59, 60–69, 70–79, and 80+), and comorbid conditions, with a normally distributed mean-zero random effect per unique patient. In this way, we compared BPs following periods of excellent vs poor adherence while controlling for measured patient characteristics and patient identity as a random effect.

Our second approach focused exclusively on 14 patients who had at least one period each of poor and excellent adherence within the 90-day study window.

TABLE I. Sample Characteristics

	Full Sample (N=869)	Analysis 1 (N=178)	Analysis 2 (N=14)
Men, %	35	29[a]	36
White, %	43	40	43
Hyperlipidemia, %	53	56	43
Diabetes, %	33	40[a]	50
Peripheral vascular disease, %	5	7	21
Renal insufficiency, %	6	6	7
Coronary artery disease, %	13	19[a]	36
Congestive heart failure, %	3	6	14
Cerebrovascular disease, %	5	4	7
Obese, %	60	60	57
Age, mean (SD), y	59.4 (11.4)	60.6 (10.7)	61.9 (12.3)
First BP reading			
SBP, mean (SD)	133.6 (17.4)	131.8 (17.5)	141.0 (18.4)
DBP, mean (SD)	80.4 (11.4)	78.7 (11.7)[a]	85.9[a] (12.2)
BP <140/90 mm Hg, %	55	58	36
Medications at baseline			
ACE inhibitor/ARB, %	66	65	57
β-Blocker, %	45	46	71
Calcium channel blocker, %	36	42	50
Diuretic, %	65	68	79
Other medication, %	12	8	14
Total number of medications, mean (SD)	2.3 (1.0)	2.4 (1.1)	2.8 (1.3)

Abbreviations: ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; BP, blood pressure; DBP, diastolic blood pressure; SBP, systolic blood pressure; SD, standard deviation. [a]$P<.05$. Significant differences at the .05 level were found between the 178 patients in analysis 1 and the 691 (869–178) distinct patients in the full sample for sex, presence of diabetes, presence of coronary artery disease, and first DBP measurement. A significant difference at the .05 level was found between the 14 patients in analysis 2 and the 164 (distinct) patients in analysis 1 for only first DBP measurement.

TABLE II. Models for Effect of Excellent vs Poor Adherence Based on 178 Patients

	Systolic Blood Pressure	Diastolic Blood Pressure
Intercept	137.55 (3.70)	89.98 (2.13)
Excellent adherence	−11.60 (2.79)[a]	−7.67 (1.61)[a]
Men	2.36 (2.61)	1.85 (1.51)
White	−2.69 (2.39)	−1.99 (1.38)
Age <60 y	Reference	Reference
Age 60–69 y	2.96 (2.66)	−3.44 (1.53)[b]
Age 70–79 y	6.37 (3.32)	−4.98 (1.92)[b]
Age 80+ y	13.47 (6.76)[b]	−10.15 (3.91)[b]
Hyperlipidemia	1.42 (2.37)	−2.39 (1.37)
Diabetes	−2.69 (2.61)	−4.15 (1.51)[b]
Peripheral vascular disease	−6.67 (4.70)	−4.20 (2.72)
Renal insufficiency	0.24 (5.12)	1.20 (2.96)
Coronary artery disease	1.06 (3.26)	−2.01 (1.88)
Congestive heart failure	2.76 (5.19)	2.24 (3.00)
Cerebrovascular disease	13.77 (5.85)[b]	0.30 (3.38)
Obese	2.35 (2.47)	1.57 (1.43)

Values are expressed as coefficient estimates (standard errors). [a]$P<.001$. [b]$P<.05$.

The analyses consisted of random effects least-squares regressions in which SBP and DBP were regressed on binary indicators of excellent vs poor adherence 7 days prior to the reading, with normally distributed mean-zero patient-specific random effects. The random effects models were fit using the ''lme'' function in the statistics software package R (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

As shown in Table I, the full cohort of 869 patients were 35% male, 43% white, and aged between 25 to 86 years with a mean age of 59.4 years (standard deviation=11.4 years). We recorded the first BP reading taken after enrollment. For 781 of the full cohort of 869 patients, this corresponded to BP taken on the day of enrollment. The average initial post-enrollment SBP and DBP were 133.6 mm Hg and 80.4 mm Hg, respectively, and 45% of the cohort had an initial BP >140/90 mm Hg. The most commonly used medication classes were angiotensin-converting enzyme inhibitors/angiotensin receptor blockers and diuretics. The average patient received 2.3 medications at baseline. For analysis 1, we studied only BP readings preceded by a 7-day period of poor (<60%) or excellent (100%) adherence, eliminating BP values preceded by a period of intermediate adherence. The sample for analysis 1 consisted of 178 patients. The characteristics of this sample were similar to the full cohort, except that the analysis 1 sample contained a significantly ($P<.05$) smaller fraction of men (29% compared with 35%), had a significantly higher proportion with diabetes (40% compared with 33%), and had a significantly higher proportion with coronary artery disease (19% compared with 13%). The average DBP was significantly lower for the analysis 1 sample compared with the full cohort (78.7 compared with 80.4).

Analysis 1 included a total of 357 BP readings for 178 unique patients. For 7-day periods with poor adherence, the average adherence rate was 34%, with a standard deviation of 22%. Table II summarizes the main results for analysis 1. Controlling for demographics and comorbid conditions, BP readings following periods of excellent adherence were lower than those following periods of poor adherence (SBP −11.6 mm Hg, DBP −7.7 mm Hg; $P<.001$ for both).

In analysis 2, 14 patients had at least 1 BP value preceded by a period of poor adherence (<60%) and at least 1 BP value preceded by a period of excellent adherence (100%). These 14 patients contributed a total of 36 observations. During the 90-day study

window, 8 of the patients had excellent adherence always followed by poor adherence, 5 had poor adherence always followed by excellent adherence, and 1 patient had both poor and excellent adherence following each other. For 7-day periods with poor adherence, the average adherence rate was 33%, with a standard deviation of 24%, nearly identical to the results based on the larger sample in analysis 1. The analysis 2 group appeared to have a higher burden of comorbidity than the analysis 1 group and higher initial BP values. However, except for a higher initial DBP in the analysis 2 group, these differences were not statistically significant. Accounting for patient random effects, mean BP following excellent adherence was 130.6/78.1 mm Hg, compared with 145.5/85.2 mm Hg following poor adherence, a difference of 14.9/7.1 mm Hg ($P<.05$ both for SBP and DBP).

In both analyses, we examined alternative window periods prior to BP readings (vs the base case definition of 7 days), and different definitions of excellent and poor adherence (vs the base case definitions of 100% and <60%). In general, the results were similar. Extending the window length excluded a larger number of patients in the excellent and poor adherence groups, and decreasing the window length resulted in somewhat less powerful effects of adherence on BP. Increasing the percent threshold for poor adherence (eg, to 80%) attenuated the effect of poor adherence on BP.

DISCUSSION

In this study, we estimated the effect size of adherence on BP control. By focusing on 7-day periods characterized by excellent (100%) or poor (<60%) adherence, we were able to show that the difference between these two is approximately 12/8 mm Hg or 15/7 mm Hg, in our first and second analyses, respectively. Our second analysis, although limited to only 14 patients, allowed each patient to serve as his or her own control. The effect size we found is robust to the method of analysis. It was similar whether we included a large number of patients and controlled for several potential confounding factors or whether we included only patients for whom both poor and excellent adherence periods were observed. These results, therefore, provide a methodologically robust estimate regarding the extent of the impact of poor adherence on BP control during a 7-day period.

Our findings have utility for clinicians who treat hypertension. Patients often arrive at a visit not having taken their medication for 1 or more days, and may communicate this fact to the clinician. We have previously shown that an impression that a patient is nonadherent is often associated with a decision not to intensify the antihypertensive regimen.[18] This relatively common occurrence (ie, admission of suboptimal adherence leading to a decision not to intensify) is likely a major contributor to clinical inertia, which, in turn, is a major barrier to improved BP control.[18–21] In another study, we have shown that, contrary to what many clinicians might expect, therapy intensification improves BP to a similar extent in patients with suboptimal adherence compared with those with optimal adherence.[22] We have therefore suggested that clinicians not dismiss the idea of intensifying therapy in a patient who is known or suspected to have suboptimal adherence.[22]

The present study adds to this line of reasoning by providing an estimate of the extent of BP elevation that can be expected after a 7-day period of poor adherence in a real-life setting. While our previous study suggested that intensification can be considered in patients with suboptimal adherence,[22] the current study suggests that intensification may be indicated when the SBP is elevated by >15 mm Hg or the DBP by >8 mm Hg. Blaming this extent of BP elevation on nonadherence may not make sense in light of the present study.

Previous efforts to estimate the effect size of nonadherence on BP have been limited by assessing both adherence and BP control in less-than-optimal ways (binary measures of control, binary measures of adherence, unclear timing between the two). In a seminal study, Morisky and colleagues[2] developed a 4-item scale to measure self-reported nonadherence and then demonstrated the criterion validity of that measure. In that study, 75% of patients deemed adherent by the scale had controlled BP at 5-year follow-up, compared with 47% of patients deemed nonadherent by the scale. In another well-known study, the authors used automated pharmacy fills data to assess adherence and again found that nonadherence during a 30-day period was a risk factor for uncontrolled BP.[3] In contrast, our study quantifies the effect size of adherence in terms of mm Hg rather than limited to a binary outcome of controlled/uncontrolled, and does so during a 7-day period. Previous studies have shown that it may not be sufficient to characterize patients as adherent or nonadherent, because patients may have periods of excellent adherence interspersed with "drug holidays," or periods during which the medication is intentionally omitted for several days.[10,23–25] Because long-term adherence is not a binary concept, it is important to understand the impact of short-term adherence on the outcome of interest rather than simply labeling some patients as nonadherent and then demonstrating that they have inferior BP control.

LIMITATIONS

While the results of our study are compelling, we do acknowledge some important limitations. First, and most importantly, we cannot establish causal effects of nonadherence from our observational data. While we controlled for important determinants of BP in our analyses, unobserved confounders could have played a role in our results. Second, we used a carefully selected subset of patients who recorded periods of excellent or poor adherence, and for the second analysis, patients

who recorded at least one of each. Not only did this sample selection necessarily limit our sample size, but it arguably could impact generalizability, although the group comparisons in Table I did not reveal large differences in the variables we measured. Still, by limiting our study to patients who had multiple BP measurements in 90 days, we may have selected for a sicker group of patients, specifically those who were more likely to have comorbidities such as diabetes and coronary heart disease. In addition, the second analysis was restricted to patients who had periods of both excellent and poor adherence, so that by design this particular sample had more erratic behavior than the general hypertensive population. However, because the estimated effect sizes from the two analyses were consistent, this concern may not be a serious one. Further, the consistent results between the two analyses greatly enhances validity because using each patient as his or her own control is arguably the gold standard for controlling for confounding due to patient-specific factors. A third limitation is that we tracked adherence using the index medication, ie, the medication whose bottle had a MEMS cap. However, most of these patients were taking other medications as well, which were not monitored by MEMS caps. This is a usual practice in adherence research,[10,17] and previous studies have shown that adherence to an index medication matches well with adherence to other medications in the regimen. Furthermore, there is no accepted method available for harmonizing the results of multiple simultaneous MEMS caps. Fourth, this study relied on actual BP measurements from clinical practice, rather than obtaining BP measurements through a standardized research protocol. While this feature of the study may enhance generalizability to real-world settings, it may also compromise the reliability of our BP data. Finally, our study enrolled patients from a single medical center, an inner-city safety-net hospital with a high proportion of minority and immigrant patients. This also may impact generalizability.

CONCLUSIONS

This study provides a fairly precise estimate of the effect of adherence on BP control in the short term: 7 days of poor adherence (<60%) increases BP by approximately 12 mm Hg to 15/7–8 mm Hg compared with 7 days of excellent adherence. Patients who admit to substantial nonadherence may nevertheless benefit from intensification of the antihypertensive regimen if their BP is elevated by more than this amount.

Acknowledgments and disclosures: Drs Rose and Glickman contributed equally to this manuscript. This research was supported by a grant from the National Institutes of Health (HL072814, NR Kressin, PI). Dr Rose is supported by a career development award from the Department of Veterans Affairs, Health Services Research and Development Service. Dr Kressin is supported in part by a Research Career Scientist award from the Department of Veterans Affairs, Health Services Research & Development Service (RCS 02-066-1). The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. The authors have no conflicts of interest to disclose.

References

1. Choo PW, Rand CS, Inui TS, et al. Validation of patient reports, automated pharmacy records, and pill counts with electronic monitoring of adherence to antihypertensive therapy. *Med Care*. 1999;37:846–857.
2. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care*. 1986;24:67–74.
3. Fung V, Huang J, Brand R, et al. Hypertension treatment in a Medicare population: adherence and systolic blood pressure control. *Clin Ther*. 2007;29:972–984.
4. Heisler M, Hogan MM, Hofer TP, et al. When more is not better: treatment intensification among hypertensive patients with poor medication adherence. *Circulation*. 2008;117:2884–2892.
5. Ho PM, Magid DJ, Shetterly SM, et al. Importance of therapy intensification and medication nonadherence for blood pressure control in patients with coronary disease. *Arch Intern Med*. 2008;168:271–276.
6. Vrijens B, Goetghebeur E. Electronic monitoring of variation in drug intakes can reduce bias and improve precision in pharmacokinetic/pharmacodynamic population studies. *Stat Med*. 2004;23:531–544.
7. Vrijens B, Goetghebeur E, de Klerk E, et al. Modelling the association between adherence and viral load in HIV-infected patients. *Stat Med*. 2005;24:2719–2731.
8. Vrijens B, Tousset E, Rode R, et al. Successful projection of the time course of drug concentration in plasma during a 1-year period from electronically compiled dosing-time data used as input to individually parameterized pharmacokinetic models. *J Clin Pharmacol*. 2005; 45:461–467.
9. Vrijens B, Urquhart J. Patient adherence to prescribed antimicrobial drug dosing regimens. *J Antimicrob Chemother*. 2005;55:616–627.
10. Vrijens B, Vincze G, Kristanto P, et al. Adherence to prescribed antihypertensive drug treatments: longitudinal study of electronically compiled dosing histories. *BMJ*. 2008;336:1114–1117.
11. Girvin BG, Johnston GD. Comparison of the effects of a 7-day period of non-compliance on blood pressure control using three different antihypertensive agents. *J Hypertens*. 2004;22:1409–1414.
12. Kressin NR, Wang F, Long J, et al. Hypertensive patients' race, health beliefs, process of care, and medication adherence. *J Gen Intern Med*. 2007;22:768–774.
13. Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med*. 2005;353:487–497.
14. Dobbels F, De Geest S, van Cleemput J, et al. Effect of late medication non-compliance on outcome after heart transplantation: a 5-year follow-up. *J Heart Lung Transplant*. 2004;23:1245–1251.
15. Kimmel SE, Chen Z, Price M, et al. The influence of patient adherence on anticoagulation control with warfarin: results from the International Normalized Ratio Adherence and Genetics (IN-RANGE) Study. *Arch Intern Med*. 2007;167:229–235.
16. Lee JY, Kusek JW, Greene PG, et al. Assessing medication adherence by pill count and electronic monitoring in the African American Study of Kidney Disease and Hypertension (AASK) Pilot Study. *Am J Hypertens*. 1996;9:719–725.
17. Qureshi NN, Hatcher J, Chaturvedi N, et al. Effect of general practitioner education on adherence to antihypertensive drugs: cluster randomised controlled trial. *BMJ*. 2007;335:1030.
18. Rose AJ, Shimada SL, Rothendler JA, et al. The accuracy of clinician perceptions of "usual" blood pressure control. *J Gen Intern Med*. 2008;23:180–183.
19. Berlowitz DR, Ash AS, Hickey EC, et al. Inadequate management of blood pressure in a hypertensive population. *N Engl J Med*. 1998;339:1957–1963.
20. Phillips LS, Branch WT, Cook CB, et al. Clinical inertia. *Ann Int Med*. 2001;135:825–834.
21. Roumie CL, Elasy TA, Wallston KA, et al. Clinical inertia: a common barrier to changing provider prescribing behavior. *Jt Comm J Qual Patient Saf*. 2007;33:277–285.
22. Rose AJ, Berlowitz DR, Manze M, et al. Intensifying therapy for hypertension despite suboptimal adherence. *Hypertension*. 2009; 54:524–529.
23. Urquhart J. Pharmacodynamics of variable patient compliance: implications for pharmaceutical value. *Adv Drug Deliv Rev*. 1998; 33:207–219.
24. Urquhart J, Vrijens B. 'Hedged' prescribing for partially compliant patients. *Clin Pharmacokinet*. 2006;45:105–108.
25. Vrijens B, Goetghebeur E. Comparing compliance patterns between randomized treatments. *Control Clin Trials*. 1997;18:187–203.

Prompt Repeat Testing After Out-of-Range INR Values
A Quality Indicator for Anticoagulation Care

Adam J. Rose, MD, MSc, FACP; Elaine M. Hylek, MD, MPH; Dan R. Berlowitz, MD, MPH; Arlene S. Ash, PhD; Joel I. Reisman, AB; Al Ozonoff, PhD

Background—Improved control of oral anticoagulation reduces adverse events. A program of quality measurement is needed for oral anticoagulation. The interval until the next test after an out-of-range International Normalized Ratio (INR) value (the "follow-up interval") could serve as a process of care measure.

Methods and Results—We studied 104 451 patients cared for by 100 anticoagulation clinics in the Veterans Health Administration (VA). For each site, we computed the average follow-up interval after low (≤1.5) or high (≥4.0) INR. Our outcome was each site's average anticoagulation control, measured by percent time in therapeutic range (TTR); 59 837 patients (57%) contributed to the low INR analysis, 37 697 (36%) contributed to the high INR analysis, and all patients contributed to the dependent variable (mean site TTR). After a low INR, site mean follow-up interval ranged from 10 to 24 days. Longer follow-up intervals were associated with worse site-level control (1.04% lower for each additional day, $P<0.001$). After a high INR, site mean follow-up interval ranged from 6 to 18 days, with longer follow-up intervals associated with worse site-level control (1.12% lower for each additional day, $P<0.001$). These relationships were somewhat attenuated but still highly statistically significant when the proportion of INR values in-range was used as the dependent variable rather than TTR.

Conclusions—Prompt repeat testing after out-of-range INR values is associated with better anticoagulation control at the site level and could be an important part of a quality improvement effort for oral anticoagulation. **(*Circ Cardiovasc Qual Outcomes*. 2011;4:276-282.)**

Key Words: anticoagulants ■ quality of health care ■ ambulatory care ■ medication therapy management ■ warfarin

Although anticoagulation therapy with warfarin is potentially life-saving, it is also potentially dangerous. Warfarin has an extremely narrow therapeutic window, and fluctuations in the degree of anticoagulation can be difficult to anticipate or prevent.[1] Meticulous control of anticoagulation, as measured by percent time in therapeutic range (TTR), has been shown to reduce the rate of adverse events in patients receiving anticoagulation, both at the level of the individual patient[2–5] and at the site of care level.[6,7] However, anticoagulation control is often suboptimal, leaving much room for improvement.[8,9] Before we can improve the quality of anticoagulation care, we must be able to measure it. Therefore, we are in need of a program of quality measurement and quality improvement in the management of oral anticoagulation. Our group has proposed an outcome measure (anticoagulation control, as measured by risk-adjusted TTR)[10] and has used it to profile 100 sites of care in the Veterans Health Administration (VA).[11]

TTR is an intermediate outcome measure that has been linked to definitive outcomes including stroke, venous thromboembolism, and major hemorrhage. Although outcome measures can be an excellent way to measure quality of care, they do not provide a prescription for action. Process measures are attractive in this regard because they do provide a prescription for action and because process can be measured at each clinical encounter.[12,13] One attractive process measure for oral anticoagulation would be promptness of repeat testing after an out-of-range International Normalized Ratio (INR) value. Although both high and low INR values have been linked to patient harm,[14–18] clinical guidelines do not give specific advice about the optimal follow-up interval after a high or low INR value.[1,19] This is probably because there have been no studies of this important issue.

We therefore used a database of 100 sites of care and over 100 000 unique patients in the VA to address 2 questions. First, do sites of care differ regarding the interval until the

The opinions expressed in this manuscript do not necessarily represent the official views of the Department of Veterans Affairs.

Received November 21, 2010; accepted February 18, 2011.

From the Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, Bedford, MA (A.J.R., D.R.B., J.I.R., A.O.); the Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA (A.J.R., E.M.H., D.R.B., A.S.A.); the Department of Health Policy and Management, Boston University School of Public Health, Boston, MA (D.R.B.); the Department of Quantitative Health Sciences, Division of Biostatistics and Health Services Research, University of Massachusetts School of Medicine, Worcester, MA (A.S.A.); and the Biostatistics Section, Boston Children's Hospital, Boston, MA (A.O.).

Correspondence to Adam J. Rose, MD, Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, 200 Springs Rd, Bldg 70, Bedford, MA 01730. E-mail adamrose@bu.edu

Circ Cardiovasc Qual Outcomes **is available at http://circoutcomes.ahajournals.org** **DOI: 10.1161/CIRCOUTCOMES.110.960096**

next INR test after a low (≤1.5) or a high (≥4.0) INR value? Second, how do these differences relate to site-level performance, as defined by risk-adjusted TTR? By demonstrating both significant variations in practice and their relationship with intermediate outcomes of care, we sought to find support for promptness of follow-up after an out-of-range INR value as a process measure for oral anticoagulation care.

WHAT IS KNOWN

- Episodes of excessive or insufficient anticoagulation increase the risks of bleeding and thromboembolism, respectively.
- There have been no previous studies regarding the ideal follow-up interval after a high (≥4) or low (≤1.5) International Normalized Ratio (INR) value.
- In the absence of empirical evidence, clinical guidelines make no specific recommendations regarding follow-up after an out-of-range INR value.

WHAT THE STUDY ADDS

- In our study, the mean interval until the next INR test after a high or low INR varied widely among 100 sites of care in an integrated health care system (from 6 to 18 days after a high INR and from 10 to 24 days after a low INR).
- Sites with shorter mean follow-up intervals had better anticoagulation control. Risk-adjusted site mean percent time in range was approximately 1% lower for each additional day of the follow-up interval after either a high or low INR.
- Follow-up within 1 week after a high or low INR appears to be ideal, based on our results; this has the potential to serve both as a performance measure and as a putative standard of care.

Methods

Data

The database for this study has also been described elsewhere.[8,11] The Veterans AffaiRs Study to Improve Anticoagulation (VARIA) included all patients deemed to be receiving oral anticoagulation therapy (OAT) from the VA between October 1, 2006, to September 30, 2008, based on the criteria described below. The study was approved by the Institutional Review Board of the Bedford VA Medical Center.

Patients

We included all patients who received warfarin from the VA during the 2-year study period (ie, at least 30 days' worth dispensed by the pharmacy) and who had at least 2 valid intervals for calculating percent TTR.[20] For this purpose, a valid interval consists of 2 INR values separated by 56 days or less, without an intervening hospitalization, as in the original study by Rosendaal et al.[20]

We excluded patients whose primary indication to receive warfarin was valvular heart disease. Many such patients have a target INR range of 2.5 to 3.5 rather than the more standard 2 to 3, but it is not possible to determine with certainty which patients have the higher target range. Without specific knowledge of the target range, we cannot calculate TTR. We also excluded patients who only recorded INR values 1.2 and lower, reasoning that most such patients received INR tests for reasons unrelated to warfarin management (eg, frequent emergency department visits).

Laboratory Values and Calculation of Percent TTR

We included INR values within the VA system that were obtained while patients were "on warfarin," that is, when a patient was either (1) in possession of warfarin or (2) having INR tests at least every 42 days. This choice of a 42-day interval is based on previous work by Go et al[21] as well as the current American College of Cardiology/American Heart Association/European Society of Cardiology guidelines for the maximum allowable follow-up interval for patients anticoagulated for atrial fibrillation.[22] We defined the period of warfarin possession as the duration of the most recent VA prescription for warfarin, plus 30 days. We excluded INR tests measured while the patient was hospitalized within the VA system. Patients who are hospitalized may receive temporary parenteral anticoagulation (eg, with heparin) or no anticoagulation; therefore out-of-range INR values while hospitalized may be intentional and do not necessarily reflect poor quality of care. For this study, we also excluded patient INR data from the first 6 months of therapy with warfarin (the "inception period"). We have previously shown that TTR is lower during the inception period,[8,11,23] and decisions regarding the follow-up interval may also differ during this period.

We calculated TTR using the Rosendaal method,[20] which uses linear interpolation to assign an INR value to each day between successive observed INR values. Gaps of 56 days or more between INR values are not interpolated. After interpolation, the percentage of time during which the interpolated INR values lie between 2.0 and 3.0 (from 0% to 100%) is calculated.[20]

Sites of Care

We included 100 VA sites of care, each of which includes a hospital, an outpatient care center, and several outlying community-based clinics. Each site has a specialized anticoagulation clinic, which is usually run by clinical pharmacists under the supervision of a medical director.[24] Therefore, essentially all patients whose anticoagulation is managed in the VA are treated by specialized anticoagulation clinics. Most patients only visited one site of care, and their INR data were assigned to that site. If a patient visited more than 1 site (3% of patients), we partitioned their data by site.

Risk Adjustment Model

We have previously described the derivation and validation of our risk adjustment model for TTR.[8] We considered many potential variables that we thought were likely to affect TTR, including demographics, area-level poverty, driving distance to care, physical health conditions, mental health conditions, number of medications, and number of hospitalizations. Most variables were retained within the model, with the exception of several comorbid conditions that did not have appreciable effect sizes. The model was derived and validated according to customary procedures, which included considerations of maximizing predictive ability, clinical credibility, and ease of use and understanding.[8] This patient-level risk adjustment model for TTR has an R^2 of 13.3% when used with this dataset.[11] Table 1 contains all the variables that were retained in the final model.

Dependent Variables: Site-Level Anticoagulation Control

For our dependent variable, we used 2 separate measures of site-level anticoagulation control. Our main dependent variable was mean site risk-adjusted TTR. We calculated risk-adjusted TTR for each patient who received anticoagulation management at our 100 sites of care, whether or not they recorded any out-of-range INR values. This is because we wanted to measure the results achieved by each site for all of its patients, as a measure of overall quality of care. Site risk-adjusted TTR was calculated using the following procedure. First, for each patient, we calculated the observed TTR ("O") and applied the risk adjustment model to calculate the expected TTR

Table 1. Baseline Sample Characteristics for the Overall Sample and Subsets of the Sample

Variable	Overall Sample (n=104 451)*	Low INR Sample (n=59 837)*	High INR Sample (n=37 697)*
Female sex	1.9%	2.2%	2.3%
Median age (IQR)	72 (62–79)	72 (61–79)	72 (61–79)
Race/ethnicity			
Non-Hispanic white	77.2%	76.6%	77.6%
Non-Hispanic black	8.5%	10.0%	9.4%
Hispanic	2.8%	3.1%	2.7%
Asian	0.3%	0.3%	0.3%
Native American	0.3%	0.3%	0.3%
Other/unknown	10.9%	9.7%	9.7%
Median % poverty in zip code of residence (IQR)	10.7 (6.6–15.9)	10.8 (6.7–16.2)	10.5 (6.5–15.8)
Median distance from nearest VA facility in miles (IQR)	7.8 (3.7–16.5)	7.5 (3.6–16.0)	7.2 (3.4–14.7)
Primary indication for warfarin†			
Atrial fibrillation	64.2%	62.9%	61.9%
Venous thromboembolism	27.3%	29.3%	30.0%
All others combined	8.5%	7.8%	8.0%
Physical comorbid conditions			
Cancer (newly diagnosed)	6.8%	8.2%	7.4%
Chronic kidney disease	14.2%	15.5%	16.3%
Chronic liver disease	1.2%	1.3%	1.5%
Chronic lung disease	29.4%	32.0%	32.4%
Diabetes mellitus	40.1%	41.7%	40.9%
Epilepsy	2.8%	3.2%	3.4%
Heart failure	32.8%	35.0%	35.6%
Hyperlipidemia	75.4%	75.6%	75.3%
Hypertension	84.0%	84.4%	84.7%
Mental comorbid conditions			
Alcohol abuse	9.3%	11.3%	12.1%
Bipolar disorder	2.3%	2.8%	2.9%
Dementia	5.3%	5.7%	5.9%
Major depression	21.6%	24.7%	24.8%
Substance abuse (nonalcohol)	4.0%	5.3%	5.2%
Median No. of medications (IQR)	8 (6–12)	9 (6–13)	9 (6–13)
Hospitalized at least once	26.2%	32.3%	32.5%
Anticoagulation control			
Percent time in range, mean (SD)	0.612 (0.219)	0.541 (0.201)	0.538 (0.180)

INR indicates International Normalized Ratio; IQR, interquartile range.

Baseline sample characteristics are for the overall sample, which was used to calculate site performance (ie, risk-adjusted percent time in range), and for subsets of the sample, which were used to characterize follow-up intervals after low or high INR values.

*All patients were included in the overall sample, which was used to measure site performance (percent time in range). The subset of patients with at least 1 low INR value (≤1.5) was included in the low INR sample, and the subset of patients with at least 1 high INR value (≥4.0) was included in the high INR sample. These groups are not mutually exclusive.

†Patients whose main indication for anticoagulation was valvular heart disease or prosthetic heart valve were excluded from this study.

("E"). Then, an observed minus expected (O-E) score was calculated for each patient. The mean O for each site constituted its unadjusted TTR, whereas the mean O-E score for each site constituted its risk-adjusted TTR. A full explanation of how we calculated risk-adjusted TTR, as well as a comparison between adjusted and unadjusted TTR, can be found in our earlier reports.[8,11]

One possible criticism of using TTR as the outcome for this study is that TTR will necessarily reward sites for following up sooner after an out-of-range value, even if the date on which the patient's control improved was in fact the same, and all that changed was the promptness of measurement. We therefore examined the extent to which our results depended on this mathematical truism. We substituted site-level proportion of INR values in range for risk-adjusted TTR as an alternate dependent variable and reran our main analyses. This allowed us to determine whether we still saw the same effect even without the measurement property related to the calculation of TTR.

Independent Variable: Site Mean Interval Until Next INR Test

We characterized each site regarding the mean interval until the next test after an out-of-range INR value (the "follow-up interval"). We divided out-of-range INR values into 2 categories: ≤1.5 ("low") and

≥4.0 ("high"). Both high and low INR values have clearly been linked to patient harm[1,14–18] and should be addressed promptly to bring the patient back within the therapeutic range. In addition, we examined follow-up intervals after INR values that are only slightly out of range: "slightly low" (1.6 to 1.9) and "slightly high" (3.1 to 3.9). The evidence regarding harm from such values is less robust, so we wanted to characterize the extent to which sites are responding more promptly to the more extreme values. In addition, we wanted to explore the effect of the management of slightly out-of-range INR values on site-level TTR.

We located all patients who had at least 1 high or low INR value (the "index value") followed by another INR within 56 days without intervening hospitalizations. For an out-of-range INR value, another INR is expected soon after; therefore, the following INR could be recorded as soon as the next day. If the patient was hospitalized between the index INR and the next value, we looked back to the next possible index value instead, because hospitalization also counts as prompt follow-up of the aberrant value. When a patient had multiple qualifying episodes of a high or low INR, we selected the last such episode, so that each patient was sampled no more than once for high and/or once for low. We also reran our analyses after selecting the first such episode or a random episode of high or low INR; the results did not change appreciably (data not shown). We averaged values from individual patients to calculate mean values for each site. We followed similar procedures to characterize each site regarding its response to slightly low (1.6 to 1.9) and slightly high (3.1 to 3.9) INR values.

Statistical Analyses

We examined the baseline characteristics of patients in our source population as well as the characteristics of patients who were included in the subsamples to analyze follow-up intervals after a high or low INR value. We characterized each site by a mean follow-up interval after a low INR (≤1.5), after a high INR (≥4.0), and after an INR value that was slightly low (1.6 to 1.9) or slightly high (3.1 to 3.9). We modeled the site-level relationships between follow-up intervals and risk-adjusted TTR using simple correlation, linear regression, and ANOVA with the Tukey honestly significant differences test (after grouping sites into quintiles by follow-up intervals). We repeated these analyses using our alternate dependent variable (proportion of INR values in range by site). All analyses were conducted using SAS, version 9.1 (SAS Corporation). Dr Rose had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Results

Patients

We studied 104 451 unique patients who received anticoagulation during the experienced period (ie, >6 months' experience with anticoagulation). Baseline characteristics for the source population are described in Table 1. The sample was mostly male (98%) and had a median age of 72 years. Most patients (64%) were anticoagulated for atrial fibrillation, with the remainder anticoagulated for venous thromboembolism (27%) or other indications (9%; eg, mural thrombus, cardiomyopathy, pulmonary hypertension, etc). The population had a substantial burden of comorbidity. For example, 40% had diabetes mellitus, 33% had heart failure, 14% had chronic kidney disease, and 7% were newly diagnosed with (nonskin) cancer during the study period. The burden of mental illness and substance abuse was also considerable: 22% had major depression, 9% had a diagnosis of alcohol abuse, and 5% had dementia.

In general, differences between the source population and the samples used to study high and low INR were slight. Some characteristics that we have previously linked to lower patient-level TTR (such as cancer, alcohol abuse, and substance abuse) were slightly more common among the high and low INR samples; these differences attained statistical significance due to the large study size (for example, $P<0.001$ for these 3 variables). Mean TTR for the source population was 61%, compared with 54% in the high and low INR groups, respectively ($P<0.001$ for both comparisons).

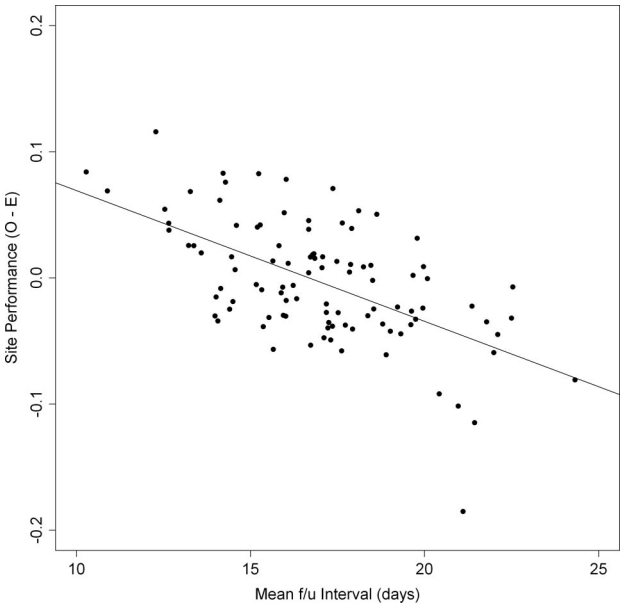


Figure 1. Correlation between mean site-level follow-up (f/u) interval after a low International Normalized Ratio value (≤1.5) and site-level performance, as measured by risk-adjusted percent time in range (n=100 sites). For the line of correlation, $r=-0.59$ ($P<0.001$), indicating that shorter follow-up intervals are associated with better site-level performance. O-E indicates observed minus expected score.

Relationship Between Follow-Up Intervals and Site-Level Anticoagulation Control

There were 100 sites of care in the database. Site mean TTR ranged from 41% to 72% and site risk-adjusted performance (site O-E score) ranged from 19% below to 12% above expected. Performance was above expected at 46 sites and below expected at 54 sites; the median site O-E score was 0.6% below expected. Sites differed widely regarding mean follow-up intervals. After a low INR (≤1.5), site mean follow-up intervals ranged from 10 to 24 days (Figure 1). After a high INR (≥4.0), site mean follow-up intervals ranged from 6 to 18 days (Figure 2). Site mean follow-up intervals after slightly low and slightly high INR ranged from 12 to 32 and from 13 to 31 days, respectively (data not shown). Generally, the sites that pursued prompter follow-up after low INR also pursued prompter follow-up after high INR. For example, the correlation between site-level mean follow-up interval after low and high INR values was 0.72.

With regard to follow-up after a low INR (≤1.5), shorter site-level intervals correlated with improved site-level performance as measured by risk-adjusted TTR (Figure 1; $r=-0.59$, $P<0.001$). This result was unchanged when unadjusted site-level TTR was used as the outcome ($r=-0.59$). With regard to follow-up after a high INR (≥4.0), shorter

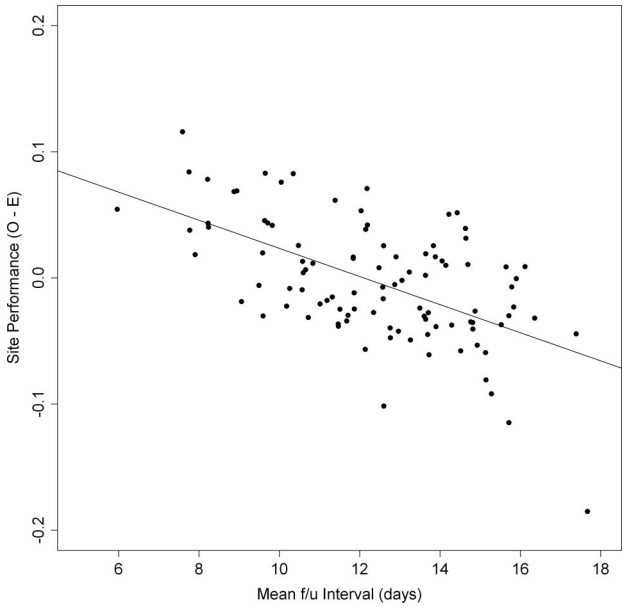


Figure 2. Correlation between mean site-level follow-up (f/u) interval after a high International Normalized Ratio value (≥4.0) and site-level performance, as measured by risk-adjusted percent time in range (n=100 sites). For the line of correlation, $r=-0.57$ ($P<0.001$), indicating that shorter follow-up intervals are associated with better site-level performance. O-E indicates observed minus expected score.

site-level intervals correlated with improved site-level performance as measured by risk-adjusted TTR (Figure 2; $r=-0.57$, $P<0.001$). This result was unchanged when unadjusted site-level TTR was used as the outcome ($r=-0.57$).

We further examined the strength of these relationships. For each additional day of mean site-level follow-up interval after a low INR, site-level risk-adjusted TTR was 1.04% lower (95% confidence interval, 0.75 to 1.32%; $P<0.001$). We found similar results when we used a cutoff of ≤1.3 to define a low INR value (data not shown). For each additional day of mean site-level follow-up interval after a high INR, site performance was 1.12% lower (95% confidence interval, 0.80 to 1.43%; $P<0.001$). We found similar results when we used a cutoff of ≥5.0 to define a high INR value (data not shown).

We repeated these analyses in subpopulations defined by indication for anticoagulation, namely patients anticoagulated for atrial fibrillation (64% of the sample) and patients anticoagulated for venous thromboembolism (27% of the sample). The main findings of the study did not change. For example, for each additional day after a low INR value, site-level risk-adjusted TTR was 1.00% lower among atrial fibrillation patients, 1.01% lower among venous thromboembolism patients, and 1.04% lower among all patients ($P<0.001$ for all 3 findings).

We also divided sites into quintiles (20 sites per group), based on the intervals after a low or high INR value. Site performance was generally best in the quintile with the shortest follow-up and worst in the quintile with the longest follow-up (Table 2), although differences among the middle 3 quintiles were small and not statistically significant.

Table 2. Quintiles of Site Follow-Up Intervals Compared With Site Performance as Measured by Risk-Adjusted Percent Time in Therapeutic Range

Quintiles of Follow-Up Interval After Index INR Value	Mean Follow-Up Interval (SD)	Mean O-E	*P* Value, ANOVA
After low INR (≤1.5)			$P<0.001$
Shortest follow-up (a)	13.4 (1.2)	3.24%	
Short (b)	15.6 (0.5)	0.77%	
Moderate (c)	16.9 (0.3)	−0.41%	
Long (d)	18.2 (0.5)	−0.76%	
Longest (e)	20.9 (1.3)	−4.49%	
After high INR (≥4.0)			$P<0.001$
Shortest follow-up (a)	8.8 (1.1)	4.20%	
Short (b)	11.1 (0.5)	−0.16%	
Moderate (c)	12.5 (0.4)	−0.62%	
Long (d)	14.0 (0.4)	−0.73%	
Longest (e)	15.6 (0.8)	−4.35%	

INR indicates International Normalized Ratio.

There are 20 sites in each quintile (total n=100 sites).

Site performance is measured by the observed minus expected (O-E) score. For each 1% increase in this score, the site performed 1% better than expected according to the risk-adjustment model.

Tukey honestly significant differences test revealed that the following groups were statistically indistinguishable at the corrected 0.05 level of significance: for low INR: a–b, b–d, and e; for high INR: a, b–d, and e.

We also examined the correlation between follow-up after mildly out-of-range INR values and site risk-adjusted TTR. These correlations were slightly less than for more pronounced deviations from the target range, but were still considerable (Table 3). For slightly low values (1.6 to 1.9), the correlation was −0.53 ($P<0.001$), compared with −0.59 for low values (≤1.5). For slightly high values (3.1 to 3.9), the correlation was −0.45 ($P<0.001$), compared with −0.57 for high values (≥4.0).

Sensitivity Analysis: Substituting Proportion of INR Values in Range for Percent Time in Range

As discussed in the methods, we substituted the proportion of INR values in range at each site for the site O-E score (ie, risk-adjusted TTR) as our dependent variable. As shown in

Table 3. Correlation Between Mean Site Follow-Up Intervals After Out-of-Range INR Values and 2 Measures of Site-Level Anticoagulation Control

	Risk-Adjusted Percent Time in Range		Proportion of INR Values in Range	
Index INR Value	*r*	*P*	*r*	*P*
≤1.5 (very low)	−0.587	<0.001	−0.415	<0.001
1.6–1.9 (somewhat low)	−0.529	<0.001	−0.286	0.004
3.1–3.9 (somewhat high)	−0.454	<0.001	−0.281	0.005
≥4.0 (very high)	−0.575	<0.001	−0.407	<0.001

INR indicates International Normalized Ratio.

Values are risk-adjusted percent time in range and proportion of INR values in range (n=100 sites).

Table 3, substitution of proportion of INR values in range lessened the magnitude of these correlations, but they were still present and statistically significant. Effect sizes decreased much more with regard to follow-up after mildly out-of-range INR values than severely out-of-range values. For example, the correlation for follow-up after low INR (≤1.5) decreased from −0.59 to −0.42 and the correlation for follow-up after slightly low INR (1.6 to 1.9) decreased from −0.53 to −0.29.

Discussion

Improving the quality of care in oral anticoagulation has the potential to save thousands of lives per year in the United States as well as preventing numerous nonfatal events that nevertheless lead to hospitalization or institutionalization.[2–7,10] To improve quality of care in oral anticoagulation, we need valid quality measures. In particular, process measures may be useful because they can provide a ready prescription for action and remediation.[10] In this study, we found that sites vary considerably with regard to promptness of repeat testing after an out-of-range INR value and that prompter repeat testing after out-of-range INR values (a process measure) is associated with improved site-level anticoagulation control (an intermediate outcome of care). This relationship was true both with and without risk adjustment for patient characteristics and both with and without linear interpolation between adjacent INR values.

Despite being part of an integrated health system, the 100 sites that we studied had a wide range of practice in this regard, possibly due to the relative lack of evidence and clear guideline recommendations. For example, after a high INR (≥4.0), site mean follow-up intervals ranged from 6 to 18 days. These variations had important consequences for anticoagulation control: Based on our regression models, a site with an average 6-day interval after a high INR would be expected to have a mean TTR 13% higher than a site with an average 18-day interval. This is a very large difference in anticoagulation control and one that has been linked to considerable differences in rates of adverse events.[2–7] It is likely that clinicians practicing at each site arrive at a consensus about the ideal follow-up interval in certain situations, whether by written policy or unwritten common practice. These site-level tendencies can presumably be changed, and our results suggest that prompter follow-up could improve TTR considerably for most of the sites in our study.

In addition to its implications for quality measurement, our study also has implications for clinical practice guidelines in anticoagulation care. Our results suggest that anticoagulation control could be improved considerably by following up within 7 days after a high (≥4.0) or low (≤1.5) INR value and within 14 days after a mildly high (3.1 to 3.9) or mildly low (1.6 to 1.9) INR value. If all VA patients had been treated in this manner during our study, our results suggest that the VA might have recorded an overall TTR between 5% to 10% higher, a difference that has been associated with meaningful improvements in the rates of outcomes such as stroke, venous thromboembolism, major hemorrhage, and mortality.[2–7]

This study has several strengths. We used a large and powerful database, rich in clinical detail. We used 2 measures of anticoagulation control, both a simple one (proportion of values in range) and one that is the result of much development by our group and represents the state of the art in quality measurement for oral anticoagulation (risk-adjusted TTR).[8,10,11] The consistency of our findings suggests that they are not attributable only to the measures used but represent a real and important finding that means exactly what one would think it means.

However, several limitations should be noted. First, we measured average follow-up intervals at the site level rather than at the level of the individual patient or the individual instance. Sites of care probably determine these follow-up intervals through written or unwritten policies, which can and should be changed to improve performance. However, follow-up for individual patients is dependent on many clinical considerations and is inherently variable. Therefore, we believe that it would be ill advised to attempt to measure quality of care for individual patients using these measures—although this issue could certainly be examined empirically. Second, we measured the actual interval that elapsed between INR values but did not measure the interval that the clinician requested. To some extent, nonadherence to recommendations on the part of the patient might have played a role in the correlation between longer follow-up intervals and poor control. However, our main outcome measure was risk-adjusted for the patient population at each site, which should have controlled for many of the patient-level factors that contribute to poor compliance and poor control. Third, this study only included patients with a target INR range of 2 to 3. Therefore, these results may not be generalizable to patients with other target ranges (most often 2.5 to 3.5). Finally, VA patients are mostly male and have a high burden of comorbidity. However, it is unclear how this fact would have altered the basic relationships that we showed between follow-up intervals and site-level anticoagulation control.

In summary, we found that there is a wide range of practice regarding the interval until a repeat test after out-of-range and mildly out-of-range INR values. Longer follow-up intervals were associated with worse anticoagulation control. We believe that we have truly identified a quality measure for oral anticoagulation care. Our study suggests that optimizing follow-up intervals after out-of-range INR values could greatly improve anticoagulation control and prevent thousands of fatal or morbid adverse events each year.

Sources of Funding

Dr Rose is supported by a career development award from the US Department of Veterans Affairs, Health Services Research, and Development Service (CDA-08-017). The sponsor had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, and approval of the manuscript.

Disclosures

Dr Hylek received honoraria from Bayer and Bristol Myers Squibb and has served on advisory boards for Boehringer-Ingelheim, Bristol Myers Squibb, Merck, and Sanofi-aventis.

References

1. Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G. Pharmacology and management of the vitamin K antagonists: American

College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest.* 2008;133:160S–198S.
2. van Leeuwen Y, Rosendaal FR, Cannegieter SC. Prediction of hemorrhagic and thrombotic events in patients with mechanical heart valve prostheses treated with oral anticoagulants. *J Thromb Haemost*. 2008;6: 451–456.
3. van Walraven C, Oake N, Wells PS, Forster AJ. Burden of potentially avoidable anticoagulant-associated hemorrhagic and thromboembolic events in the elderly. *Chest*. 2007;131:1508–1515.
4. Veeger NJ, Piersma-Wichers M, Tijssen JG, Hillege HL, van der Meer J. Individual time within target range in patients treated with vitamin K antagonists: main determinant of quality of anticoagulation and predictor of clinical outcome. A retrospective study of 2300 consecutive patients with venous thromboembolism. *Br J Haematol*. 2005;128:513–519.
5. White HD, Gruber M, Feyzi J, Kaatz S, Tse HF, Husted S, Albers GW. Comparison of outcomes among patients randomized to warfarin therapy according to anticoagulant control: results from SPORTIF III and V. *Arch Intern Med*. 2007;167:239–245.
6. Connolly SJ, Pogue J, Eikelboom J, Flaker G, Commerford P, Franzosi MG, Healey JS, Yusuf S. Benefit of oral anticoagulant over antiplatelet therapy in atrial fibrillation depends on the quality of international normalized ratio control achieved by centers and countries as measured by time in therapeutic range. *Circulation*. 2008;118:2029–2037.
7. Wallentin L, Yusuf S, Ezekowitz MD, Alings M, Flather M, Franzosi MG, Pais P, Dans A, Eikelboom J, Oldgren J, Pogue J, Reilly PA, Yang S, Connolly SJ. Efficacy and safety of dabigatran compared with warfarin at different levels of international normalised ratio control for stroke prevention in atrial fibrillation: an analysis of the RE-LY trial. *Lancet*. 2010;376:975–983.
8. Rose AJ, Hylek EM, Ozonoff A, Ash AS, Reisman JI, Berlowitz DR. Patient characteristics associated with oral anticoagulation control: results of the Veterans AffaiRs Study to Improve Anticoagulation (VARIA). *J Thromb Haemost*. 2010;8:2182–2191.
9. van Walraven C, Jennings A, Oake N, Fergusson D, Forster AJ. Effect of study setting on anticoagulation control: a systematic review and metaregression. *Chest*. 2006;129:1155–1166.
10. Rose AJ, Berlowitz DR, Frayne SM, Hylek EM. Measuring quality of oral anticoagulation care: extending quality measurement to a new field. *Jt Comm J Qual Patient Saf*. 2009;35:146–155.
11. Rose AJ, Hylek EM, Ozonoff A, Ash AS, Reisman JI, Berlowitz DR. Risk-adjusted percent time in therapeutic range as a quality indicator for outpatient oral anticoagulation: results of the Veterans AffaiRs study to Improve Anticoagulation (VARIA). *Circ Cardiovasc Qual Outcomes*. 2011;4:22–29.
12. Brook RH, McGlynn EA, Cleary PD. Quality of health care, part 2: measuring quality of care *N Engl J Med*. 1996;335:966–970.
13. Donabedian A. Explorations in quality assessment and monitoring. Ann Arbor, MI: Health Administration Press; 1980.
14. Hylek EM, Go AS, Chang Y, Jensvold NG, Henault LE, Selby JV, Singer DE. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. *N Engl J Med*. 2003;349:1019–1026.
15. Hylek EM, Singer DE. Risk factors for intracranial hemorrhage in outpatients taking warfarin. *Ann Intern Med*. 1994;120:897–902.
16. Hylek EM, Skates SJ, Sheehan MA, Singer DE. An analysis of the lowest effective intensity of prophylactic anticoagulation for patients with nonrheumatic atrial fibrillation. *N Engl J Med*. 1996;335:540–546.
17. Linkins LA, Choi PT, Douketis JD. Clinical impact of bleeding in patients taking oral anticoagulant therapy for venous thromboembolism: a meta-analysis. *Ann Intern Med*. 2003;139:893–900.
18. Rose AJ, Ozonoff A, Grant RG, Henault LE, Hylek EM. The epidemiology of sub-therapeutic anticoagulation in the United States. *Circ Cardiovasc Qual Outcomes*. 2009;2:591–597.
19. Ansell JE, Oertel LB, Wittkowsky AK, eds. *Managing Oral Anticoagulation Therapy: Clinical and Operational Guidelines*. 3rd ed: Wolters-Kluwer Health: Philadelphia, PA: 2009.
20. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briet E. A method to determine the optimal intensity of oral anticoagulant therapy. *Thromb Haemost*. 1993;69:236–239.
21. Go AS, Hylek EM, Chang Y, Phillips KA, Henault LE, Capra AM, Jensvold NG, Selby JV, Singer DE. Anticoagulation therapy for stroke prevention in atrial fibrillation: how well do randomized trials translate into clinical practice? *JAMA*. 2003;290:2685–2692.
22. Fuster V, Ryden LE, Cannom DS, Crijns HJ, Curtis AB, Ellenbogen KA, Halperin JL, Le Heuzey JY, Kay GN, Lowe JE, Olsson SB, Prystowsky EN, Tamargo JL, Wann S, Smith SC Jr, Jacobs AK, Adams CD, Anderson JL, Antman EM, Halperin JL, Hunt SA, Nishimura R, Ornato JP, Page RL, Riegel B, Priori SG, Blanc JJ, Budaj A, Camm AJ, Dean V, Deckers JW, Despres C, Dickstein K, Lekakis J, McGregor K, Metra M, Morais J, Osterspey A, Tamargo JL, Zamorano JL. ACC/AHA/ESC 2006 Guidelines for the management of patients with atrial fibrillation. *Circulation*. 2006;114:e257–e354.
23. Rose AJ, Ozonoff A, Henault LE, Hylek EM. Warfarin for atrial fibrillation in community-based practise. *J Thromb Haemost*. 2008;6: 1647–1654.
24. Department of Veterans Affairs: Veterans Health Administration. VHA Directive 2009-003: Anticoagulation Therapy Management. Washington, DC: February 2, 2009.

Risk-Adjusted Percent Time in Therapeutic Range as a Quality Indicator for Outpatient Oral Anticoagulation

Results of the Veterans Affairs Study To Improve Anticoagulation (VARIA)

Adam J. Rose, MD, MSc; Elaine M. Hylek, MD, MPH; Al Ozonoff, PhD; Arlene S. Ash, PhD; Joel I. Reisman, AB; Dan R. Berlowitz, MD, MPH

Background—Oral anticoagulation is safer and more effective when patients receive high-quality care. However, there have been no prior efforts to measure quality of oral anticoagulation care or to risk adjust it to ensure credible comparisons. Our objective was to profile site performance in the Veterans Health Administration (VA) using risk-adjusted percent time in therapeutic range (TTR).

Methods and Results—We included 124 551 patients who received outpatient oral anticoagulation from 100 VA sites of care for indications other than valvular heart disease from October 1, 2006, to September 30, 2008. We calculated TTR for each patient and mean TTR for each site of care. Expected TTR was calculated for each patient and each site based on the variables in the risk adjustment model, which included demographics, comorbid conditions, medications, and hospitalizations. Mean TTR for the entire sample was 58%. Site-observed TTR varied from 38% to 69% or from poor to excellent. Site-expected TTR varied from 54% to 62%. Site risk-adjusted performance ranged from 18% below expected to 12% above expected. Risk adjustment did not alter performance rankings for many sites, but for other sites, it made an important difference. For example, the site ranked 27th of 100 before risk adjustment was one of the best (risk-adjusted rank, 7). Risk-adjusted site rankings were consistent from year to year (correlation between years, 0.89).

Conclusions—Risk-adjusted TTR can be used to profile the quality of outpatient oral anticoagulation in a large, integrated health system. This measure can serve as the basis for quality measurement and quality improvement efforts. **(*Circ Cardiovasc Qual Outcomes*. 2011;4:22-29.)**

Key Words: anticoagulants ■ quality of health care ■ ambulatory care ■ risk adjustment ■ patients ■ safety

Oral anticoagulation is a highly effective but potentially dangerous therapy.[1,2] The level of anticoagulation control is a critical determinant of benefit from warfarin[3–7]; indeed, patients with atrial fibrillation may not benefit from anticoagulation unless they achieve a certain level of control.[3] However, warfarin management is difficult, and achieving good control requires much effort and skill on the part of both the patient and the clinician.[1] Because many patients do not achieve excellent control,[8] there is great potential to improve outcomes for patients by improving quality of care in oral anticoagulation.[9] To improve quality of care, we first must be able to measure it.[10,11] Previous efforts to measure quality of care in oral anticoagulation therapy have focused disproportionately on the failure to provide anticoagulation to as many ideal candidates as possible.[12] However, receipt of anticoagulation is only a first step toward improving outcomes; we also need to measure the quality of oral anticoagulation management to ensure that the benefits of anticoagulation are maximized and the harms minimized.[9]

An ideal quality indicator for outpatient oral anticoagulation would have several characteristics: it would be easy to abstract, calculate, and understand; it would vary among providers or sites of care; improvement would be possible; and there would be strong evidence linking it to important outcomes, such as stroke, venous thromboembolism, and major hemorrhage. Percent time in therapeutic range (TTR) has many of these characteristics. It can be calculated from automated data, it can be improved,[8] and it has been linked to important outcomes.[3–7] Although it might be possible to consider using definitive outcomes themselves as quality indicators, reliance on these rare events would preclude quality measurement at all but the largest sites of care.[9,13]

Received May 24, 2010; accepted October 6, 2010.

From the Center for Health Quality, Outcomes, and Economic Research (A.J.R., A.O., J.I.R., D.R.B.), Bedford VA Medical Center, Bedford, Mass; Department of Medicine (A.J.R., E.M.H., A.S.A., D.R.B.), Section of General Internal Medicine, Boston University School of Medicine, Boston, Mass; Biostatistics Section (A.O.), Boston Children's Hospital, Boston, Mass; Department of Quantitative Health Sciences (A.S.A.), Division of Biostatistics and Health Services Research, University of Massachusetts School of Medicine, Worcester, Mass; and Department of Health Policy and Management (D.R.B.), Boston University School of Public Health, Boston, Mass.

The opinions expressed in this manuscript do not necessarily represent the official views of the Department of Veterans Affairs.

The online-only Data Supplement is available at http://circoutcomes.ahajournals.org/cgi/content/full/CIRCOUTCOMES.110.957738/DC1.

Correspondence to Adam J. Rose, MD, MSc, Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, 200 Springs Rd, Building 70, Bedford, MA 01730. E-mail adamrose@bu.edu

Circ Cardiovasc Qual Outcomes **is available at http://circoutcomes.ahajournals.org** **DOI: 10.1161/CIRCOUTCOMES.110.957738**

One reason why TTR has not been used previously as a quality indicator may be the absence of a risk adjustment model for TTR. Risk adjustment can enhance the credibility of performance comparisons between providers or sites by ensuring that sites are being compared regarding quality of care rather than regarding merely differences in case mix.[14] Risk adjustment can increase the acceptance of quality measures by poorly performing sites who might otherwise protest that their performance is poor because their patients are sicker. Our group recently has derived and validated a risk adjustment model for TTR that should allow for fair site-site comparisons on this quality indicator.[15] In deriving and validating our model, we confirmed the widely held belief that some patients are indeed much harder to keep within the target range than others, suggesting that risk adjusting TTR is necessary.

We therefore set out to address 3 related questions, using a database of 100 anticoagulation clinics and >100 000 patients from the Veterans Health Administration (VA). First, does mean TTR differ among sites of care? Unless meaningful differences exist, profiling is unlikely to spur quality improvement. Second, does risk adjusting TTR meaningfully alter site rankings? Risk adjustment may enhance credibility,[14] but it requires effort, so it is important to know whether it matters. Finally, would risk-adjusted site rankings be relatively constant from year to year, suggesting that risk-adjusted TTR is measuring quality of care (a stable attribute of a site) rather than measuring mere statistical variation? Our overarching objective was to examine the suitability of risk-adjusted TTR as a potential quality indicator.

WHAT IS KNOWN

- The safety and effectiveness of oral anticoagulation can be improved by better control, (ie, more time in therapeutic range [TTR]).
- Although oral anticoagulation is prescribed for millions of patients each year, there has been no organized approach to measuring or improving the quality of oral anticoagulation.

WHAT THE STUDY ADDS

- We used clinic-level risk-adjusted TTR to profile the performance of 100 anticoagulation clinics in an integrated system of care (the Veterans Health Administration).
- We propose the use of risk-adjusted TTR as a quality indicator to measure and track the quality of oral anticoagulation in the Veterans Health Administration and other integrated health systems.
- Quality measurement in oral anticoagulation is a necessary prerequisite to quality improvement, which holds the promise of preventing adverse events due to inadequate or excessive anticoagulation.

Methods

Patients

The database for this study also has been described elsewhere.[15] The VA is the largest integrated health system in the United States, and for many years, has collected comprehensive data regarding the care delivered to its patients, including inpatient care, outpatient care, and pharmacy records. The Veterans Affairs Study to Improve Anticoagulation included all patients deemed to be receiving oral anticoagulation therapy from the VA between October 1, 2006, and September 30, 2008, based on the criteria described later. The study was approved by the Institutional Review Board of the Bedford VA Medical Center.

A flowchart of study inclusion criteria is shown in Figure 1. We included all patients who received warfarin from the VA during the 2-year study period (ie, at least 30 days' worth dispensed by the pharmacy) and who had at least 2 valid intervals for calculating TTR.[16] For this purpose, a valid interval consists of 2 international normalized ratio (INR) values separated by ≤56 days without an intervening hospitalization.

There were 2 levels of exclusions: individual patients and entire sites of care. On the patient level, we excluded patients whose primary indication to receive warfarin was valvular heart disease. Many such patients have a target INR range of 2.5 to 3.5 rather than the more standard 2 to 3, but it is not possible to determine with certainty which patients have the higher target range. Without specific knowledge of the target range, we cannot calculate TTR. We also excluded patients whose only recorded INR values were ≤1.2, reasoning that most such patients received INR tests for reasons unrelated to warfarin management (eg, frequent emergency department visits).

Sites of Care

There are 128 sites of care within the VA, each of which includes a hospital, an outpatient care center, and several outlying community-based clinics. Each site has a specialized anticoagulation clinic, which usually is run by clinical pharmacists under the supervision of a medical director.[17] Therefore, essentially all patients whose anticoagulation is managed in the VA are managed by specialized anticoagulation clinics. Self-testing devices are not covered by the VA; therefore, few patients have or use them.

We excluded 28 sites from our study and several months of data from an additional 14 sites because our data-checking procedures revealed possible problems with data completeness at those sites. The problem with data completeness relates to the laboratory data only. Although accurate data are collected about which laboratory tests are drawn (because something akin to a billing code is generated), the data regarding laboratory results must be checked carefully. Specifically, the name given to each laboratory test by the local facility is not uniform throughout the system, and these names may change unexpectedly; after this happens, there may be a period of several months where the local laboratory results are not captured by the national database until the name change is noted. We identified which sites had this issue by dividing the data into 3-month periods; problematic sites had few or no INR results in certain periods, whereas the number of INR tests drawn remained constant over time. In contrast, there were 86 sites that had complete data for the entire 2-year study period and 14 sites that began to have complete data during the period (usually early in the study) and continued to have it through to the end. Thus, 28 sites were excluded because of incomplete data, and 14 sites were partially included. The 14 partially included sites performed similarly to the 86 sites with complete data, suggesting that they differed only in data collection rather than in performance (results not shown). Most patients only visited 1 site of care, and their INR data were assigned to that site. If a patient visited >1 site (3% of patients), we partitioned his or her data by site.

Laboratory Values and Calculation of TTR

We included INR values within the VA system when patients were on warfarin, that is, when a patient was either (1) in possession of warfarin or (2) having INR tests at least every 42 days. We defined the period of warfarin possession as the duration of the most recent VA prescription for warfarin plus 30 days. Because patients may be instructed to take half-doses of warfarin, we recognize that going >30 days beyond the end of a prescription does not necessarily

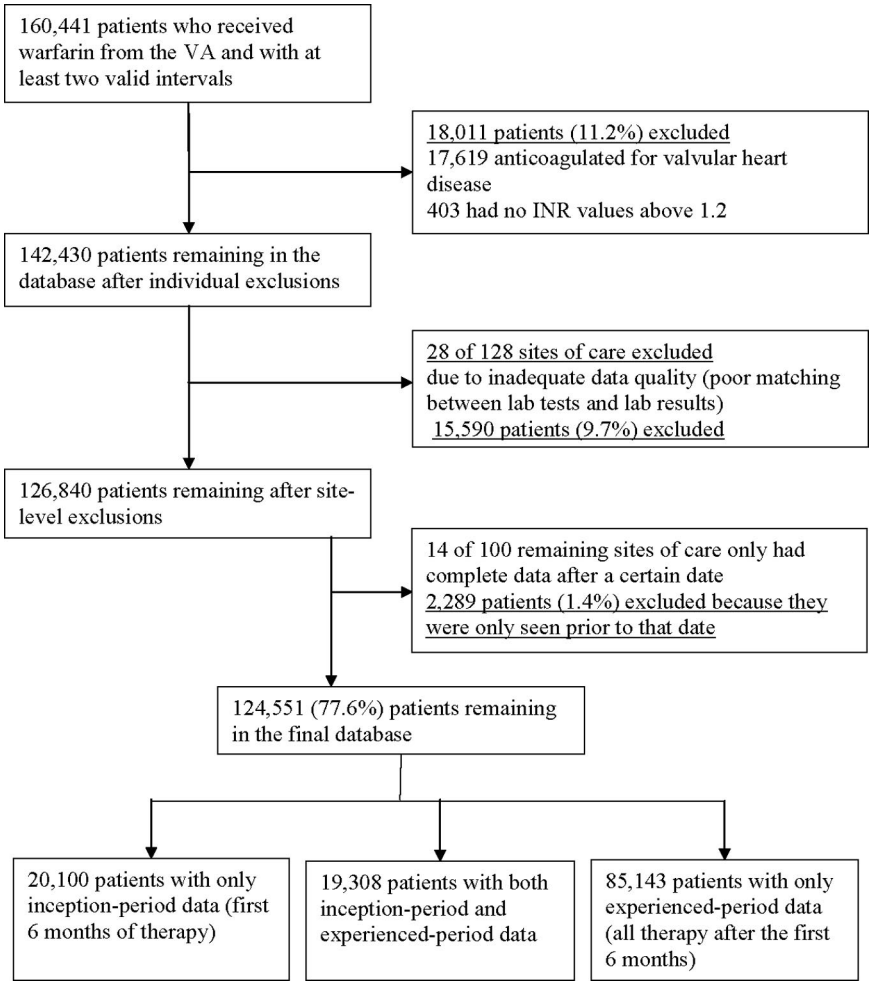


Figure 1. Flowchart of included and excluded patients.

indicate that warfarin therapy has stopped. Therefore, we also allowed a consistent pattern of INR measurements (ie, every 42 days or fewer) to indicate that a patient was still being managed. A similar approach was used to define time on warfarin in a previous study.[18]

We excluded INR tests measured while the patient was hospitalized within the VA system. Patients who are hospitalized may receive temporary parenteral anticoagulation (eg, with heparin) or no anticoagulation, so out-of-range INR values while hospitalized may be intentional and do not necessarily reflect poor quality of care.

We calculated TTR using the Rosendaal method,[16] which uses linear interpolation to assign an INR value to each day between successive observed INR values. Gaps of ≥56 days between INR values are not interpolated. After interpolation, the percentage of time during which the interpolated INR values lie between 2.0 and 3.0 (from 0% to 100%) is calculated.[16]

Risk Adjustment Model

We have previously described the derivation and validation of our risk adjustment model for TTR.[15] We considered many potential variables that we believed were likely to affect TTR, including demographics, area-level poverty, driving distance to care, physical health conditions, mental health conditions, number of medications, and number of hospitalizations. Most variables were retained within the model, with the exception of several comorbid conditions that did not have appreciable effect sizes. The model was derived and validated according to customary procedures, which included considerations of maximizing predictive ability, clinical credibility, and ease of use and understanding.[15] The Table contains all the variables that were retained in the final model.

Inception and Experienced Management of Warfarin

We calculated TTR and built separate risk adjustment models for 2 time periods—the first 6 months of therapy (inception period) and all care after 6 months (experienced period)—because these 2 periods are qualitatively different with regard to anticoagulation management. We defined each patient's date of warfarin inception, looking back as far as October 1, 2005. Inception was defined as the first INR value >1.2 or the first outpatient warfarin fill, whichever came first. It would be extremely unusual for a patient to record an INR value >1.2 unless he or she had taken warfarin. We then divided the sample into inception time (the first 6 months of warfarin therapy for each patient) and experienced time (any time thereafter). A single patient might contribute only to the inception data set (if he or she had <6 months of therapy), only to the experienced data set (if he or she began warfarin >6 months before the inception of our study), or both. Coefficients for the 2 models differed substantially,[15] confirming that the 2 periods should be risk adjusted separately.

Statistical Analyses

First, we calculated TTR for each patient during the inception and experienced periods. We then calculated a mean observed TTR (O) for each site and period. We applied our risk adjustment models to calculate the expected (E) TTR first for each patient and then for each site on the basis of the patient population at that site. We then calculated an observed minus expected (O-E) score for each site for the 2-year study period. Each site had 2 O-E scores: 1 for inception management and 1 for experienced management. Site O-E scores for the 2 time periods were highly correlated, so we combined them into

Table. Baseline Sample Characteristics for the Overall Sample, the Inception Cohort (ie, First 6 Months of Anticoagulation Therapy), and the Experienced Cohort (All Management Thereafter)

Variable	Overall Sample (n=124 551)	Inception Cohort (n=39 408)	Experienced Cohort (n=104 451)
Female sex	2589 (2.1)	1046 (2.7)	1983 (1.9)
Age, y	72 (61–79)	66 (59–76)	72 (62–79)
Race/ethnicity			
Non-Hispanic white	95 312 (76.5)	29 106 (73.9)	80 682 (77.2)
Non-Hispanic black	11 240 (9.0)	4455 (11.3)	8847 (8.5)
Hispanic	3570 (2.9)	1148 (2.9)	2976 (2.8)
Asian	368 (0.3)	134 (0.3)	302 (0.3)
Native American	360 (0.3)	148 (0.4)	279 (0.3)
Other/unknown	13 701 (11.0)	4417 (11.2)	11 365 (10.9)
Poverty in zip code of residence, %	10.7 (6.6–16.0)	10.9 (6.7–16.4)	10.7 (6.6–15.9)
Distance from nearest VA facility, miles	7.8 (3.7–16.5)	7.6 (3.6–15.9)	7.8 (3.7–16.5)
Primary indication for warfarin*			
Atrial fibrillation	76 894 (61.7)	21 568 (54.7)	67 045 (64.2)
Venous thromboembolism	36 402 (29.2)	13 933 (35.4)	28 567 (27.3)
All others combined	11 255 (9.0)	3907 (9.9)	8839 (8.5)
Physical comorbid conditions			
Cancer (newly diagnosed)	9236 (7.4)	3938 (10.0)	7091 (6.8)
Chronic kidney disease	17 333 (13.9)	5224 (13.3)	14 804 (14.2)
Chronic liver disease	1584 (1.3)	565 (1.4)	1253 (1.2)
Chronic lung disease	36 138 (29.0)	11 006 (27.9)	30 666 (29.4)
Diabetes	48 874 (39.2)	14 420 (36.6)	41 842 (40.1)
Epilepsy	3485 (2.8)	1090 (2.8)	2926 (2.8)
Heart failure	38 860 (31.2)	10 129 (25.7)	34 208 (32.8)
Hyperlipidemia	91 844 (73.7)	26 957 (68.4)	78 711 (75.4)
Hypertension	103 300 (82.9)	31 343 (79.5)	87 733 (84.0)
Mental comorbid conditions			
Alcohol abuse	12 695 (10.2)	5338 (13.5)	9719 (9.3)
Bipolar disorder	3109 (2.5)	1278 (3.2)	2382 (2.3)
Dementia	6356 (5.1)	1723 (4.4)	5511 (5.3)
Major depression	27 432 (22.0)	9380 (23.8)	22 569 (21.6)
Substance abuse (nonalcohol)	5827 (4.7)	2673 (6.8)	4228 (4.0)
No. of medications	8 (6–12)	…	…
Hospitalized at least once	34 328 (27.6)	…	…

Data are presented as no. (%) or median (interquartile range).

*Patients whose main indication for anticoagulation was valvular heart disease or prosthetic heart valve were excluded from this study.

a single O-E score for each site, weighting by the number of patients in each period.

A site was considered a high or low outlier performer if its combined O-E score deviated from 0 by ≥5% and if the O-E score differed from 0 at the 0.01 level of significance (using *Z* scores). We used the 0.01 level of significance as opposed to the traditional 0.05 to address issues of multiple testing. A 5% absolute difference in TTR has been cited by other studies as constituting a meaningful difference in performance and is probably the agreed-on standard for a clinically important difference.[19]

We also investigated whether empirical Bayesian methods would change determinations of outlier status.[20] This technique has been shown to alter results regarding outlier status, particularly when sample sizes are small.[21] Typically, empirical Bayesian estimation shrinks estimates for the observed site mean TTR toward the overall mean, and fewer sites are outliers when analyzed by such methods. However, we observed almost no shrinkage presumably because of the large sample sizes even at the smaller sites of care. Therefore, we concluded that Bayesian methods were not necessary for this situation and relied on *Z* scores to determine outliers.

We compared site rankings before and after risk adjustment to determine its effect. We calculated the proportion of the variance in TTR that was explained by patient characteristics, site-level variability, and both together. We also divided the study period in half (ie, fiscal year 2007 [FY07] versus FY08) to evaluate the stability of site O-E scores and rankings between the 2 years. Analyses were conducted using SAS version 9.1 and R version 2.8 statistical software. Dr Rose had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Results

Patient Population

We studied 124 551 unique patients who received anticoagulation from 100 sites of care (Figure 1). Baseline characteristics for the sample are described in the Table. The sample was mostly men (98%) with an average age of 72 years. Most (62%) patients were anticoagulated for atrial fibrillation, with the remainder anticoagulated for venous thromboembolism (29%) or other indications (9%) (eg, mural thrombus, cardiomyopathy, pulmonary hypertension). The population had a substantial burden of comorbidity, for example, 39% had diabetes, 31% had heart failure, 14% had chronic kidney disease, and 7% had received a new diagnosis of (nonskin) cancer during the study period. The burden of mental illness and substance abuse also was considerable: 22% had major depression, 10% had received a diagnosis of alcohol abuse, and 5% had dementia. As would be expected with a population of sicker patients, they received many medications (median of 8), and 28% were hospitalized at least once during the 2-year study period.

Site Performance

The median number of patients per site was 1584 (range, 103 to 5103; interquartile range, 1043 to 2345). Mean site-expected TTR during inception was based on patient population at each site and varied from 44% to 50%. Mean site-expected TTR during the experienced period varied from 58% to 65%. We examined case mix at the 5 sites with the lowest predicted TTR (difficult case mix) and the 5 sites with the highest predicted TTR (easy case mix). The difficult sites treated more minority patients and more patients living in high-poverty zip codes. Patients at the difficult sites had higher rates of almost every comorbid condition that is associated with lower TTR in our database,[15] often several-fold higher. Finally, patients at the difficult sites received more medications and were more likely to be hospitalized. In short, there was no single factor that explained these differences in case mix; all the variables appeared to play a role. These between-site differences in case mix will be addressed more fully in a separate article.

During the inception period, the mean observed TTR was 48%; site mean observed TTR varied from 30% to 59% during inception. Site O-E scores for inception, which measured the difference between observed and expected TTR, ranged from −17% to +13%. Of the 100 sites of care, during the inception period, there were 12 high outliers, indicating performance at least 5% better than expected and O-E different from 0 at the 0.01 level of significance. There were 12 low outliers, indicating performance at least 5% worse than expected and O-E different from 0 at the 0.01 level of significance. During the experienced period, the mean observed TTR was 61%; site mean observed TTR varied from 41% to 72%. Site O-E scores for the experienced period ranged from −19% to +12%. Of the 100 sites of care, during the experienced period, there were 14 high outliers and 10 low outliers.

The correlation between site O-E scores for the inception and experienced periods was high (r=0.89; P<0.001) (Figure 2), suggesting that they measure a similar construct and may be combined. Mean combined TTR for the entire sample was 58%, and sites varied from 38% to 69%. Site-expected TTR ranged from 54% to 62%, and O-E scores ranged from −18% to +12%. Of the 100 sites of care, there were 13 high outliers and 10 low outliers. The 100 sites of care in our study are shown in Figure 3; a tabular form of these results appears in the online-only data supplement. The overall performance of our risk adjustment model was as follows: Patient-level characteristics alone explained 13.3% of the variability in TTR, site-level variability alone explained 2.9%, and both together explained 15.9%.

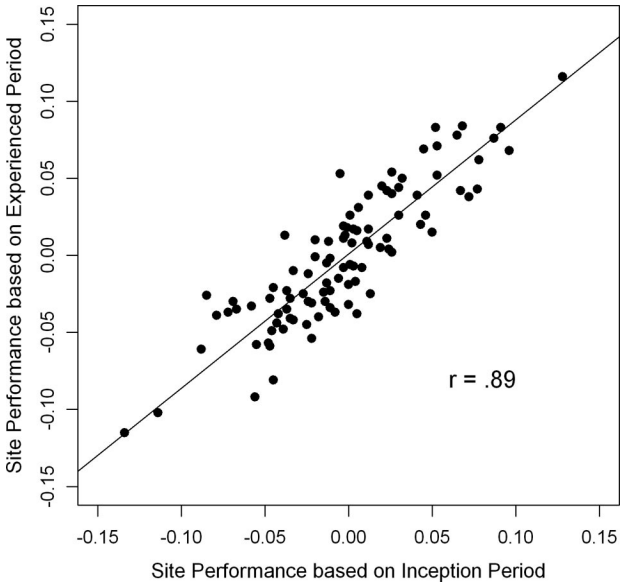


Figure 2. Comparison of site performance between the inception period (first 6 months of therapy) and the experienced period (all therapy thereafter). Performance is measured by the O-E score, which compares the TTR achieved by each site to the TTR that would be expected based on its patient population. Possible O-E scores range from −1 to +1, with positive scores indicating better-than-expected performance. Performance in the 2 periods was highly correlated (r=0.89; P<0.001).

Risk adjustment changed site rankings only slightly; there was a high degree of correlation between site rankings before and after risk adjustment (r=0.93; P<0.001). However, risk adjustment made an important difference for several sites of care. For example, site FC was ranked 27th of 100 before risk adjustment (observed TTR, 60.6%, or 2.7% above average). However, site FC had a challenging patient population (expected TTR, 53.5%, or 4.4% below average). Therefore, after risk adjustment, site FC was revealed to be one of the best performers (O-E, +7.0%; adjusted rank, 7th) as well as a high outlier. In contrast, site GI was ranked 95th before risk adjustment (observed TTR, 50.2%, or 7.7% below average) and would have been a low outlier using a definition based on unadjusted results. However, this site also had a challenging patient population (expected TTR, 53.9%, or 4% below average); although the risk-adjusted performance was poor (O-E, −3.7%; adjusted rank, 80th), it did not meet our definition for an outlier site. Not all sites were helped by risk adjustment, of course; for some sites, it invited a harsher view of their performance in light of their relatively easy case mix.

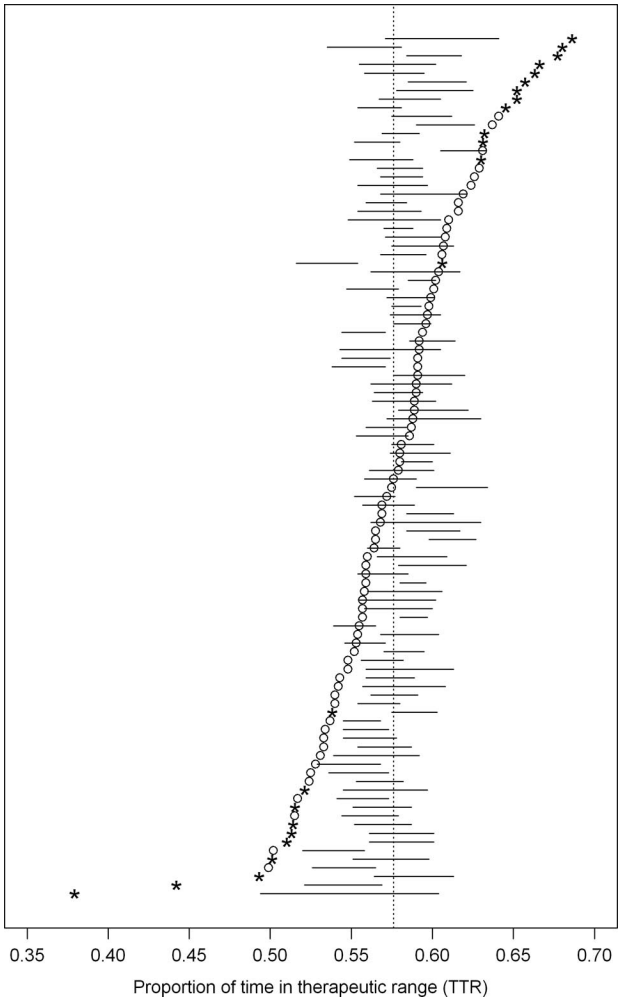


Figure 3. Overall site performance as measured by observed mean site TTR minus expected mean site TTR. The vertical line indicates the overall mean TTR (≈58% time in range). Each horizontal bar indicates the 99% CI around a site's expected mean TTR, whereas each ○ indicates the observed mean site TTR. For outlier sites (observed TTR differs from expected by ≥5%), the ○ is replaced by an *.

For example, site JC was ranked 79th before risk adjustment (observed TTR, 53.8%, or 4.1% below average), but because of its relatively easy case mix (expected TTR, 58.9%), the site was an outlier (O-E, −5.0%; adjusted rank, 92nd).

Stability of Rankings Between Years

We also divided our study period in half and compared the 2 years to each other. Most (62%) patients had INR values in both years, but 18% were only followed in FY07 and 20% only in FY08. We examined the stability of site performance between the 2 years. There was a high correlation between site performance in FY07 and FY08 both by O-E score and by risk-adjusted site rankings (r=0.89 and 0.88, respectively; P<0.001 for both) (Figure 4). Additionally, outliers were relatively consistent between years. For example, of the 9 high outliers in FY07 (year 1 of the study), 8 sites were also high outliers in FY08, whereas of the 8 sites that were low outliers in FY07, 6 were also low outliers in FY08.

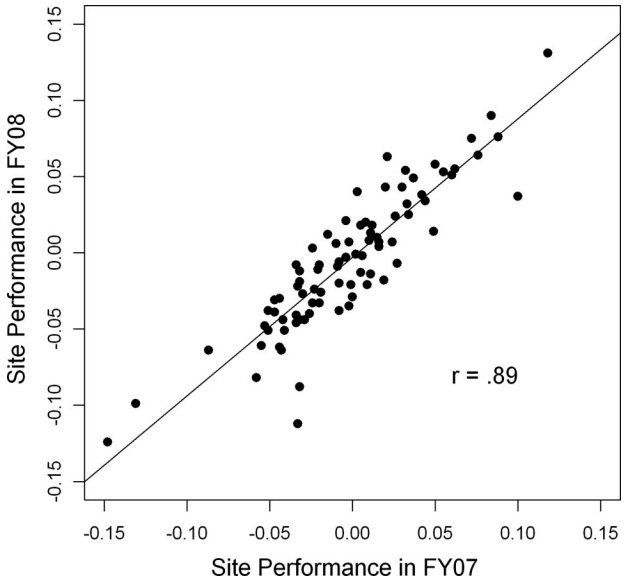


Figure 4. Comparison of site performance between FY07 and FY08. Performance is measured by the O-E score, which compares the TTR achieved by each site to the TTR that would be expected based on its patient population. Possible O-E scores range from −1 to +1, with positive scores indicating better-than-expected performance. Performance in the 2 years was highly correlated (r=0.89; P<0.001).

Discussion

In this study, we demonstrate that risk-adjusted TTR can be used as a quality indicator for oral anticoagulation care. Within the VA system, risk-adjusted TTR varies widely, with some sites performing as much as 18% below or 12% above expected. Risk-adjusted TTR is feasible to measure and is relatively consistent from year to year, suggesting that it is measuring an aspect of quality of care that is stable over time. This measure could be used by the VA or other integrated systems of care to profile annual performance and serve as an aid and impetus for quality improvement.

Risk adjustment is important for enhancing the credibility of site profiling; without risk adjustment, sites could claim that their poor performance was solely because of their case mix.[14] However, we found that the range of case mix was much smaller than the range of observed performance and that sites with very difficult or very easy patient populations were found equally among the best- and worst-performing sites (Figure 3), suggesting that although case mix varies among VA sites, the quality of oral anticoagulation care delivered to those patients varies even more widely.

We observed wide variations in performance within the VA system, from very-low TTR (38%) to very high (69%). However, after risk adjustment, site-site differences only accounted for 2.9% of the variability in TTR. Although this finding may sound like a small proportion of variability to explain, it is actually quite similar to the findings of previous studies. For example, using a population of patients with diabetes, Hofer et al[13] found that after risk adjustment, ≤4% (and as little as 1%) of the variability in hospitalization rates, visit rates, laboratory utilization rates, and glycemic control was attributable to differences between physicians. In discussing this result, these authors suggested that standardizing

processes of care among providers may not suffice to produce the quality improvement we seek; it may also be necessary to improve process of care on a system level. We therefore plan to use the present study as a basis on which to build a comprehensive program of quality measurement and quality improvement in anticoagulation care.[9] Our goal would be to improve the mean TTR in the VA to at least 70% (from the current 58%). Achieving this goal will require a focus on system-level approaches to improve processes of care. For example, our group[22] has shown that more judicious decisions regarding when to change the dose of warfarin can improve TTR considerably. Addressing this and other processes of care in a systematic way could greatly improve TTR in the VA; however, continually measuring risk-adjusted TTR will be a precondition to any program of quality improvement.[9]

Future studies also could expand this work outside the VA setting. A system of quality measurement, such as the one described in this article, would be relatively easy to implement in any integrated health system with a well-developed electronic medical record. However, the majority of patients are cared for in smaller community practices, not integrated health systems. The benefits of quality measurement could be extended to even more patients by incorporating risk-adjusted TTR into existing, commercially available anticoagulation management software, allowing sites to not only track their performance, but also potentially compare it to other sites after adjustment for case mix.

This study has important strengths, including the fact that we profiled performance among 100 sites of care within the nation's largest integrated healthcare system. In addition, the careful construction of the risk adjustment model and the considerable detail contained within its predictor variables improve our confidence that we have adjusted for case mix and that the residual variability in site performance is attributable to the quality of care.

As with any study, we acknowledge limitations. Our risk adjustment model did not include some factors that contribute to variability in TTR, especially diet and adherence to therapy. However, it could also be argued that dietary variation and medication nonadherence potentially are remediable, so adjusting for these factors could erase an important opportunity for sites to improve TTR. Second, our analyses combined patients with different indications for therapy, especially atrial fibrillation and venous thromboembolism. Although our risk adjustment model included a variable for indication, we were concerned that these groups might be too different to be combined in 1 model. To address this concern, we ran our analyses separately for patients with each indication; site performance by O-E scores did not change appreciably. Third, because of uncertainty regarding the target INR range, our study did not include patients anticoagulated for valvular heart disease, including prosthetic heart valves. However, we have every reason to expect that sites that provide superior care for patients with other indications for therapy also would provide superior care for patients with valvular heart disease. Fourth, because of strict requirements for completeness of data, we were unable to profile performance at some sites of care. Fifth, it is likely that some of the patients in our inception cohort had previously received warfarin outside the VA. If these patients were excluded from the inception cohort, we might have observed lower inception TTR; however, this is unlikely to affect site-site comparisons. Finally, although we examined TTR as a quality indicator, other summary measures of INR control might be used. For example, proportion of INR values in range could allow for more-frequent feedback about performance than TTR, which can only be calculated over a several-month period. Monthly proportion of values in range could therefore be used to provide real-time performance feedback, enhancing quality improvement efforts. In the future, we plan to compare TTR with other measures.

In summary, this article establishes that risk-adjusted TTR can be used to profile anticoagulation care in a large integrated healthcare system. Patients deserve to receive the very best care, particularly when the stakes are as high as they are with warfarin, but our report demonstrates that some patients are receiving much better care than others. It is time to shine the light of performance measurement on oral anticoagulation care.

Acknowledgments

We thank Barbara G. Bokhour, PhD, for her contributions to this manuscript.

Sources of Funding

This study was funded by the US Department of Veterans Affairs, Health Services Research and Development Service (Career Development Award to Dr Rose). The sponsor had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, and approval of the manuscript.

Disclosures

Dr Rose is supported by a Career Development Award from the Veterans Affairs Health Services Research and Development Service. Dr Hylek has received honoraria from Bayer and Bristol Myers Squibb and has served on advisory boards for Boehringer Ingelheim, Bristol Myers Squibb, Merck, and Sanofi Aventis.

References

1. Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G. Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th edition). *Chest.* 2008;133:160S–198S.
2. Budnitz DS, Shehab N, Kegler SR, Richards CL. Medication use leading to emergency department visits for adverse drug events in older adults. *Ann Intern Med*. 2007;147:755–765.
3. Connolly SJ, Pogue J, Eikelboom J, Flaker G, Commerford P, Franzosi MG, Healey JS, Yusuf S. Benefit of oral anticoagulant over antiplatelet therapy in atrial fibrillation depends on the quality of international normalized ratio control achieved by centers and countries as measured by time in therapeutic range. *Circulation*. 2008;118:2029–2037.
4. van Leeuwen Y, Rosendaal FR, Cannegieter SC. Prediction of hemorrhagic and thrombotic events in patients with mechanical heart valve prostheses treated with oral anticoagulants. *J Thromb Haemost*. 2008;6:451–456.
5. van Walraven C, Oake N, Wells PS, Forster AJ. Burden of potentially avoidable anticoagulant-associated hemorrhagic and thromboembolic events in the elderly. *Chest*. 2007;131:1508–1515.
6. White HD, Gruber M, Feyzi J, Kaatz S, Tse HF, Husted S, Albers GW. Comparison of outcomes among patients randomized to warfarin therapy according to anticoagulant control: results from SPORTIF III and V. *Arch Intern Med*. 2007;167:239–245.
7. Veeger NJ, Piersma-Wichers M, Tijssen JG, Hillege HL, van der Meer J. Individual time within target range in patients treated with vitamin K antagonists: main determinant of quality of anticoagulation and predictor

of clinical outcome. A retrospective study of 2300 consecutive patients with venous thromboembolism. *Br J Haematol*. 2005;128:513–519.
8. van Walraven C, Jennings A, Oake N, Fergusson D, Forster AJ. Effect of study setting on anticoagulation control: a systematic review and metaregression. *Chest*. 2006;129:1155–1166.
9. Rose AJ, Berlowitz DR, Frayne SM, Hylek EM. Measuring quality of oral anticoagulation care: extending quality measurement to a new field. *Jt Comm J Qual Patient Saf*. 2009;35:146–155.
10. Brook RH, McGlynn EA, Cleary PD. Quality of health care. Part 2: measuring quality of care. *N Engl J Med*. 1996;335:966–970.
11. Donabedian A. *Explorations in Quality Assessment and Monitoring.* Ann Arbor, MI: Health Administration Press; 1980.
12. Ogilvie IM, Newton N, Welner SA, Cowell W, Lip GY. Underuse of oral anticoagulants in atrial fibrillation: a systematic review. *Am J Med.* 2010;123:638–645, e634.
13. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *JAMA*. 1999;281:2098–2105.
14. Iezzoni LI, ed. *Risk Adjustment for Measuring Health Care Outcomes.* 3rd ed. Chicago, IL: Health Administration Press; 2003.
15. Rose AJ, Hylek EM, Ozonoff A, Ash AS, Reisman JI, Berlowitz DR. Patient characteristics associated with oral anticoagulation control: results of the Veterans AffaiRs Study to Improve Anticoagulation (VARIA). *J Thromb Haemost*. 2010;8:2182–2191.
16. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briet E. A method to determine the optimal intensity of oral anticoagulant therapy. *Thromb Haemost*. 1993;69:236–239.
17. Department of Veterans Affairs, Veterans Health Administration. VHA Directive 2009–003. "Anticoagulation Therapy Management." February 2, 2009.
18. Go AS, Hylek EM, Chang Y, Phillips KA, Henault LE, Capra AM, Jensvold NG, Selby JV, Singer DE. Anticoagulation therapy for stroke prevention in atrial fibrillation: how well do randomized trials translate into clinical practice? *JAMA*. 2003;290:2685–2692.
19. Samsa GP, Matchar DB. Relationship between test frequency and outcomes of anticoagulation: a literature review and commentary with implications for the design of randomized trials of patient self-management. *J Thromb Thrombolysis*. 2000;9:283–292.
20. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* West Sussex, UK: Wiley; 2004.
21. Berlowitz DR, Christiansen CL, Brandeis GH, Ash AS, Kader B, Morris JN, Moskowitz MA. Profiling nursing homes using Bayesian hierarchical modeling. *J Am Geriatr Soc*. 2002;50:1126–1130.
22. Rose AJ, Ozonoff A, Berlowitz DR, Henault LE, Hylek EM. Warfarin dose management affects INR control. *J Thromb Haemost*. 2009;7:94–101.

SUPPLEMENTAL MATERIAL

Site Rankings Based on Risk-Adjusted Percent Time in Range (TTR)

		TTR scores			rank of TTR score (lower = better)		z-score		99% CI Around Expected		
station	N	O	E	O - E	O	O - E	of O - E	Outlier?	lo 99%	E	hi 99%
IY	256	0.686	0.606	0.079	1	4	5.86	*	0.571	0.606	0.641
JW	604	0.680	0.558	0.122	2	1	13.88	*	0.535	0.558	0.581
IE	1,063	0.677	0.601	0.076	3	6	11.44	*	0.584	0.601	0.618
IW	573	0.666	0.579	0.088	4	2	9.70	*	0.555	0.579	0.602
EO	891	0.663	0.577	0.086	5	3	11.90	*	0.558	0.577	0.595
DF	934	0.657	0.603	0.053	6	11	7.52	*	0.585	0.603	0.621
FL	574	0.652	0.602	0.050	7	13	5.53	*	0.578	0.602	0.625
ET	875	0.652	0.586	0.066	8	8	8.99	*	0.567	0.586	0.605
FI	1,711	0.645	0.568	0.077	9	5	14.68	*	0.554	0.568	0.581
NK	905	0.641	0.594	0.047	10	16	6.47		0.575	0.594	0.612
EW	975	0.637	0.608	0.029	11	26	4.20		0.590	0.608	0.626
GM	2,369	0.632	0.581	0.051	12	12	11.41	*	0.569	0.581	0.592
KC	1,643	0.631	0.566	0.065	13	9	12.26	*	0.552	0.566	0.580
ID	1,585	0.631	0.619	0.012	14	37	2.15		0.605	0.619	0.633
EY	808	0.630	0.568	0.062	15	10	8.12	*	0.549	0.568	0.588
FA	1,681	0.629	0.580	0.049	16	15	9.23		0.566	0.580	0.594
JA	1,908	0.626	0.581	0.045	17	17	9.10		0.568	0.581	0.594
NJ	697	0.624	0.575	0.049	18	14	5.97		0.554	0.575	0.597
KN	434	0.619	0.594	0.024	19	27	2.34		0.568	0.594	0.621
EE	1,902	0.616	0.571	0.045	20	18	8.98		0.559	0.571	0.584
JH	809	0.616	0.573	0.043	21	19	5.59		0.554	0.573	0.593
HE	383	0.610	0.576	0.034	22	23	3.08		0.548	0.576	0.605
EX	3,677	0.609	0.579	0.031	23	25	8.58		0.570	0.579	0.588
GH	1,082	0.608	0.588	0.020	24	29	3.03		0.571	0.588	0.605
OL	857	0.607	0.594	0.013	25	35	1.73		0.575	0.594	0.613
JT	1,591	0.606	0.582	0.024	26	28	4.34		0.568	0.582	0.596
FC	861	0.606	0.535	0.070	27	7	9.54	*	0.516	0.535	0.554
DE	401	0.604	0.590	0.015	28	33	1.36		0.562	0.589	0.617
HV	3,918	0.602	0.594	0.009	29	39	2.57		0.585	0.594	0.602
HO	1,205	0.601	0.563	0.038	30	20	6.09		0.547	0.563	0.579
FG	1,571	0.599	0.587	0.012	31	36	2.19		0.572	0.587	0.601
FQ	3,746	0.598	0.584	0.013	32	34	3.79		0.575	0.584	0.593
OI	1,293	0.597	0.589	0.007	33	42	1.23		0.574	0.589	0.605
FE	2,361	0.596	0.588	0.008	34	40	1.82		0.576	0.588	0.599
EJ	1,724	0.594	0.558	0.036	35	22	6.84		0.544	0.558	0.571
DV	1,529	0.592	0.600	-0.008	36	53	-1.39		0.586	0.600	0.614
EF	324	0.592	0.574	0.018	37	30	1.48		0.543	0.574	0.605
EK	1,441	0.591	0.559	0.032	38	24	5.60		0.544	0.559	0.574
FK	1,159	0.591	0.554	0.037	39	21	5.76		0.538	0.554	0.571
EC	636	0.591	0.598	-0.007	40	52	-0.85		0.576	0.598	0.620
MB	498	0.590	0.587	0.003	41	44	0.33		0.562	0.587	0.612
JQ	1,384	0.590	0.579	0.010	42	38	1.76		0.564	0.579	0.594
KO	824	0.589	0.582	0.006	43	43	0.81		0.563	0.582	0.602
GW	691	0.589	0.600	-0.012	44	56	-1.45		0.579	0.600	0.622
JZ	375	0.588	0.601	-0.014	45	58	-1.22		0.572	0.601	0.630
EA	1,942	0.587	0.571	0.015	46	32	3.11		0.559	0.571	0.584
JP	1,166	0.586	0.569	0.017	47	31	2.74		0.553	0.569	0.585
IA	1,946	0.581	0.588	-0.007	48	51	-1.44		0.575	0.588	0.601
HN	904	0.580	0.593	-0.013	49	57	-1.76		0.574	0.593	0.611
KG	3,526	0.580	0.591	-0.011	50	55	-3.03		0.581	0.591	0.600
FJ	755	0.579	0.581	-0.002	51	47	-0.24		0.561	0.581	0.601
EN	1,258	0.576	0.574	0.002	52	46	0.26		0.558	0.574	0.590
FU	650	0.575	0.612	-0.037	53	81	-4.39		0.590	0.612	0.634
GG	1,914	0.572	0.564	0.008	54	41	1.54		0.552	0.564	0.577
JK	1,198	0.569	0.573	-0.004	55	48	-0.57		0.557	0.573	0.589
NI	1,464	0.569	0.598	-0.030	56	71	-5.23		0.584	0.598	0.613
KI	266	0.568	0.596	-0.028	57	67	-2.07		0.562	0.596	0.630
FM	1,162	0.565	0.601	-0.035	58	78	-5.58		0.584	0.601	0.617
JE	1,468	0.565	0.612	-0.048	59	90	-8.41		0.598	0.612	0.627
JF	3,110	0.564	0.570	-0.006	60	50	-1.57		0.560	0.570	0.580
IM	669	0.560	0.588	-0.028	61	68	-3.37		0.566	0.588	0.609
EG	1,249	0.559	0.569	-0.011	63	54	-1.72		0.554	0.569	0.585
KF	698	0.559	0.600	-0.041	62	86	-5.03		0.579	0.600	0.621
JY	5,113	0.559	0.588	-0.030	64	70	-9.77		0.580	0.588	0.596
EV	584	0.558	0.583	-0.025	65	64	-2.84		0.560	0.583	0.606

LY	560	0.557	0.578	-0.021	66	61	-2.29		0.555	0.578	0.602
GN	721	0.557	0.579	-0.022	67	63	-2.71		0.558	0.579	0.600
IX	4,307	0.557	0.588	-0.032	68	75	-9.61		0.580	0.588	0.597
DY	1,846	0.555	0.552	0.003	69	45	0.60		0.539	0.552	0.565
EQ	942	0.554	0.586	-0.032	70	76	-4.50		0.568	0.586	0.604
KS	1,983	0.553	0.559	-0.005	71	49	-1.11		0.546	0.558	0.571
DH	1,981	0.552	0.582	-0.030	72	73	-6.18		0.570	0.582	0.595
DD	1,845	0.548	0.569	-0.021	73	62	-4.18		0.556	0.569	0.582
EM	435	0.548	0.586	-0.038	74	82	-3.67		0.559	0.586	0.613
HB	1,473	0.543	0.574	-0.031	75	74	-5.46		0.559	0.574	0.589
KJ	489	0.542	0.582	-0.041	76	85	-4.14		0.557	0.582	0.608
IG	1,467	0.540	0.577	-0.037	77	79	-6.48		0.562	0.577	0.591
DX	1,860	0.540	0.567	-0.027	78	66	-5.36		0.554	0.567	0.580
JC	1,553	0.538	0.589	-0.050	79	92	-9.17	*	0.575	0.589	0.603
HI	2,356	0.537	0.557	-0.020	80	60	-4.57		0.545	0.557	0.568
IQ	1,596	0.534	0.559	-0.026	81	65	-4.71		0.545	0.559	0.573
ES	1,135	0.533	0.562	-0.029	82	69	-4.53		0.545	0.562	0.578
FR	1,160	0.533	0.571	-0.038	83	83	-6.03		0.554	0.571	0.587
EZ	436	0.531	0.565	-0.034	84	77	-3.32		0.539	0.565	0.592
EI	837	0.528	0.548	-0.020	85	59	-2.68		0.529	0.548	0.568
IU	903	0.525	0.555	-0.030	86	72	-4.14		0.536	0.555	0.573
FP	1,458	0.524	0.567	-0.043	87	87	-7.66		0.553	0.567	0.582
MU	459	0.521	0.571	-0.050	88	91	-4.93	*	0.545	0.571	0.597
HG	1,223	0.517	0.557	-0.040	89	84	-6.44		0.541	0.557	0.573
EU	981	0.515	0.569	-0.054	90	93	-7.85	*	0.551	0.569	0.587
GO	1,049	0.515	0.561	-0.047	91	89	-7.01		0.544	0.561	0.579
GD	1,010	0.514	0.570	-0.056	92	94	-8.17	*	0.552	0.570	0.587
GE	769	0.513	0.581	-0.069	93	95	-8.80	*	0.561	0.581	0.601
KM	763	0.510	0.581	-0.072	94	96	-9.12	*	0.561	0.581	0.601
GI	833	0.502	0.539	-0.037	95	80	-4.95		0.520	0.539	0.558
IS	584	0.501	0.575	-0.074	96	97	-8.26	*	0.551	0.575	0.598
KT	814	0.499	0.546	-0.046	97	88	-6.10		0.526	0.545	0.565
BA	504	0.493	0.588	-0.095	98	98	-9.89	*	0.564	0.588	0.613
JR	539	0.442	0.545	-0.104	99	99	-11.10	*	0.521	0.545	0.569
BB	103	0.379	0.549	-0.170	100	100	-7.97	*	0.494	0.549	0.604

O = Observed TTR; E = Expected TTR; O-E = Observed Minus Expected TTR

An outlier is a site whose observed TTR differs from its expected TTR by at least 5% and for which the difference is significant at the 0.01 level.

Relevance of Current Guidelines for Organizing an Anticoagulation Clinic

Adam J. Rose, MD, MSc; Elaine M. Hylek, MD, MPH; Al Ozonoff, PhD; Arlene S. Ash, PhD; Joel I. Reisman, AB; Patricia P. Callahan, RPh; Margaret M. Gordon, PharmD; and Dan R. Berlowitz, MD, MPH

Objective: To describe variations in the structure of anticoagulation clinic (ACC) care within the Veterans Health Administration (VA) and to identify structures of care that are associated with better site-level anticoagulation control.

Study Design: Questionnaire correlated with automated clinical data.

Methods: We characterized 90 VA ACCs using a questionnaire administered by the VA Central Office. Site descriptors included staffing levels, provider training, visit modalities, quality improvement programs, documentation, and care coordination. Patient outcomes were measured by site mean risk-adjusted percentage time in therapeutic range, a measure of anticoagulation control over time. Our study was powered to detect a 3% difference in risk-adjusted percentage time in therapeutic range, a small-to-moderate effect size, between sites with and without a certain characteristic.

Results: We observed considerable variation in the structure of ACC care. For example, 48 sites had fewer than 400 patients per provider, 25 sites had 400 to 599 patients per provider, and 17 sites had 600 patients or more per provider. However, none of the site characteristics measured were significantly related to anticoagulation control.

Conclusions: We found substantial variation in guideline-targeted organizational and management features of ACC care within the VA. However, no single feature was associated with better anticoagulation control. Current guidelines for organizing an ACC may have limited relevance for improving patient outcomes.

(Am J Manag Care. 2011;17(4):284-289)

For author information and disclosures, see end of text.

Millions of patients receive oral anticoagulation therapy to treat or prevent thromboembolic disease.[1] While effective management of warfarin sodium therapy is not easy, better anticoagulation control can improve outcomes and reduce adverse events.[2] Indeed, a growing body of literature has shown that patients whose care is managed in dedicated anticoagulation clinics (ACCs) have better outcomes than those managed in usual care.[3-6] Clinical guidelines[1,7] recommend that all ACCs should have certain features, including quality improvement programs, adequate documentation of care, sufficient training for providers, and a ratio of patients to providers of less than 400:1. However, there is little evidence as to which, if any, of these features contribute to the improved outcomes seen with ACC care.

Although oral anticoagulation therapy has been used for more than 5 decades, quality measurement has had little penetration into this field.[2] Our group recently undertook the first systematic effort to profile risk-adjusted anticoagulation control in an integrated health system. Within the Veterans Health Administration (VA), a system in which essentially all anticoagulation therapy is managed in an ACC, the mean percentage time in therapeutic range (TTR) varied widely among sites,[8] with some sites performing more than 10% better or worse than would be expected based on the risk-adjustment model. Variations of this magnitude are associated with important differences in rates of stroke, venous thromboembolism, and major hemorrhage.[9-13] To help all sites in the VA improve their performance, it is necessary to understand the site-level correlates of better or worse anticoagulation control.

Therefore, this study had the following 2 objectives: first, to describe differences in the organization and management of ACCs within the VA, and second, to assess whether these differences help explain site-level differences in anticoagulation control. We measured site characteristics using a survey of VA ACCs that examined structural features that seem likely to affect anticoagulation control, including staffing ratios, provider training protocols, and the existence of quality improvement programs.[1,7] We hypothesized that these variables would be associated with anticoagulation control. In this study, either a positive finding or a negative finding would be useful because it can help determine whether

In this article
Take-Away Points / p285
www.ajmc.com
Full text and PDF
Web exclusive
eAppendix

enforcing conformity to any specific feature is likely to improve anticoagulation care and control in the VA.

Take-Away Points

Dedicated anticoagulation clinics (ACCs) improve anticoagulation control and patient outcomes.

- Clinical guidelines recommend that ACCs should have certain features, but these recommendations are based on expert opinion rather than on empirical evidence of benefit.
- In our study, 90 ACCs in an integrated system of care differed widely in their organization and management; however, none of these differences were consistently related to ACC performance as measured by anticoagulation control, an important intermediate outcome of anticoagulation care.
- Current guidelines for organizing and managing an ACC may have limited relevance for improving outcomes.

METHODS

Data

The data for this study have been described elsewhere. The Veterans Affairs Study to Improve Anticoagulation[8,14] included all patients who received oral anticoagulation therapy from the VA between October 1, 2006, and September 30, 2008. The study was approved by the institutional review board of the Bedford VA Medical Center, Bedford, Massachusetts.

We included international normalized ratio (INR) values when patients were "on warfarin" (ie, when a patient was in possession of warfarin or was having INR tests performed at least every 42 days). We defined the period of warfarin possession as the duration of the most recent VA prescription for warfarin, plus 30 days. We calculated percentage TTR using the method by Rosendaal et al,[15] which assigns an INR value to each day by linear interpolation between successively observed INR values. Gaps of 56 days or more between INR values are not interpolated. The patient's TTR equals the percentage of days for which the interpolated INR values lie between 2.0 and 3.0 (from 0%-100%).

Independent Variables of Site-Level Characteristics

There are 128 sites of care within the VA, each of which includes a hospital, an outpatient care center, and several outlying community-based clinics. Each site has a specialized ACC, which is usually run by clinical pharmacists under the supervision of a medical director.[16] Therefore, essentially all anticoagulation care within the VA is delivered within specialized ACCs. In October 2006, VA Pharmacy Benefits Management surveyed all 128 VA sites of care for the organization and management of their ACCs. Topics included visit modalities (face to face, telephone, and mail), quality improvement programs, clinic staffing, provider training, documentation, and care coordination (the verbatim text of the questionnaire is given in the eAppendix [available at **www.ajmc.com**]). We abstracted our independent variables from the responses to the questionnaire.

Dependent Variable of Risk-Adjusted Anticoagulation Control

Site mean TTR was adjusted for case mix using a model that incorporates patient demographics, comorbid conditions that have an adverse effect on TTR, and general measures of comorbidity, including the number of medications and the number of hospitalizations. The derivation and validation of this model, which achieved an R^2 of 13.3%, has been described previously.[8,14] The model was used to calculate the expected mean TTR for each site ("E") based on patient characteristics. The expected mean TTR was compared with the observed mean TTR for each site ("O"). Therefore, each site's performance was characterized by an O minus E ("O − E") score, our measure of risk-adjusted TTR (RA-TTR).

Statistical Analysis and Power

We began by first calculating the mean observed TTR and the mean expected TTR for each site of care and then computing an O − E score for each site. We compared O − E scores between sites with and without various organizational characteristics using unpaired *t* test and compared O − E scores for multilevel variables using analysis of variance. We did not adjust for multiple comparisons. With regard to statistical power, for a characteristic present at half of the sites, *t* test would have 80% power to detect a 3% difference in the O − E score (a small-to-moderate effect size). Analyses were performed using commercially available statistical software (SAS version 9.1; SAS Institute, Cary, North Carolina).

RESULTS

Study Sample

Of 128 sites in the VA, 28 were excluded because their data were insufficiently complete to fully assess TTR. Of the remaining 100 sites, 5 did not respond to the questionnaire. Five other sites included more than 1 ACC, and the responses we received indicated that practices at the multiple ACCs of these 5 sites were not uniform. However, the dependent variable (RA-TTR) was assessed at the level of the overall site. Because we were unable to match our data for structure and outcomes, we excluded these 5 sites as well. This left 90 sites with complete data on structure and outcomes of care. The mean (SD) number of patients managed at each site was 1244 (799). The site mean TTR ranged from 38% to 69% (median,

Table 1. Binary (Yes or No) Characteristics of 90 Veterans Health Administration (VA) Anticoagulation Clinics and Relationship With Risk-Adjusted Anticoagulation Control as Measured by the Mean Percentage Time in Therapeutic Range

Site Characteristic	No. of Sites	Performance Difference[a]	*P*
Visit format			
>50% Of visits face to face	50	−0.6	.56
Any telephone visits	82	2.9	.10
Any mail notification of normal results	31	0.4	.73
Quality improvement programs			
Tracking of out-of-range INRs	48	−1.4	.17
Tracking of patients lost to follow-up	35	−0.4	.67
Formal plan-do-check-act cycles	7	−1.0	.61
Any of the above	59	−0.9	.38
Support and academic affiliation			
Clerical support	43	−1.0	.30
Pharmacy residents	44	−0.2	.87
Any support vs none	77	1.0	.46
Documentation and patient tracking			
Paper	14	−0.6	.69
VA EMR	57	1.0	.32
No. of community-based outpatient clinics under management			
≥1	61	−0.4	.73
≥3	34	−0.3	.75
Coordination of care			
Documentation of non-VA laboratory values in EMR	17	−0.7	.59
Allows monitoring by a non-VA provider	17	−0.1	.97
Formal training protocol for new providers	41	−0.6	.56

EMR indicates electronic medical record; INRs, international normalized ratios.
[a]The performance difference is positive when a characteristic predicts better anticoagulation control; a difference of 1 denotes 1% more time spent in therapeutic range.

58%). Site O − E scores ranged from −17% (ie, 17% below the expected value) to 12%.

Structure of Care and Relationship With Performance

We observed considerable variation in structure of care (Table 1). About half of the sites (n = 50) conducted most of their visits face to face; at other sites, most care was provided by telephone or mail. Most sites (n = 59) had some sort of quality improvement program, although only 7 sites used formal plan-do-check-act cycles. Most sites (n = 77) had some sort of support; 43 sites had clerical support. About half of the sites (n = 41) had a formal protocol for training new clinic providers. Many sites (n = 48) adhered to the recommended staffing ratio of less than 400 patients per provider (Table 2), although 17 sites had 600 or more patients per provider. Sites differed markedly for the perceived likelihood of being informed about a new drug-drug interaction.

Despite these differences in structure of care, we found no statistically significant predictors of site-level performance. A finding of marginal statistical significance was that 8 sites that did not allow telephone follow-up visits had almost 3% worse TTR (P = .10). There was also a hint, although not statistically significant, that sites with fewer than 500 patients under management might have worse control (performance difference, −1.6%).

Sensitivity Analyses

We performed sensitivity analyses by altering the format of the dependent variable. To ensure that our risk-adjustment model was not obscuring relationships between structure and outcomes, we also examined the same independent variables

as predictors of the mean unadjusted TTR by site, with similar results (data not shown). We also tried categorizing site O − E scores into the top quintile, bottom quintile, and all others and repeated our analyses; the results were unchanged (data not shown).

We also performed sensitivity analyses by altering the format of the independent variables. For example, we varied the cutoffs for what constituted a small site of care or for what constituted adequate staffing levels; the results were unchanged. Finally, we created a combination quality score by assigning a point for each putative quality indicator that a site fulfilled, to examine whether these factors predicted performance better in the aggregate than individually. For example, a site might receive 1 point for having a dedicated program to train providers, 1 point for having fewer than 400 patients per provider, and 1 point for using a computerized system for documentation and patient tracking. We examined several different versions of this combination quality score, including various measures and assigning them variable relative weights, but none of these combination scores predicted site RA-TTR.

Table 2. Categorical Characteristics of 90 Veterans Health Administration Anticoagulation Clinics and Relationship With Risk-Adjusted Anticoagulation Control as Measured by the Mean Percentage Time in Therapeutic Range

Site Characteristic	No. of Sites	Performance Difference[a]	*P*[b]
Patients under management			.59
<500	11	−1.6	
500-999	32	-0.1	
≥1000	47	—	
Patients per provider			.73
<400	48	—	
400-599	25	−0.3	
≥600	17	−1.1	
Visits face to face, %			.70
0-34	34	0.4	
35-74	21	0.7	
≥75	35	—	
Perceived likelihood of being informed about new drug-drug interaction, %			.94
≤25	44	0.2	
50	12	−0.7	
75	7	−0.4	
>75	25	—	

[a]The performance difference is positive when a characteristic predicts better anticoagulation control; a difference of 1 denotes 1% more time in therapeutic range.
[b]Analysis of variance.

DISCUSSION

We examined site-level organizational factors (structure of care) as potential predictors of outpatient oral anticoagulation control (outcomes of care).[17] Although our hypotheses were supported by consensus guidelines,[1,7] the site-level characteristics that we studied were not associated with anticoagulation control. Many studies relating structure and processes of care have been performed within the VA among sample sizes similar to that of this study. One such study[18] found that organizational culture and commitment to continuous quality improvement were not associated with better processes of care for depression. However, most other studies have found associations between various aspects of organizational culture and processes of care, including rates of cancer screening[19-22] and recommended processes of care for heart failure,[23] chronic lung disease,[24] and diabetes mellitus.[25] In the diabetes study, the authors also found associations between organizational culture and intermediate outcomes of diabetes care,[25] a relationship more directly comparable to our examination of structure and intermediate outcomes. However, unlike these previous studies, we directly measured structural aspects of care within the ACC itself, which we expected to more directly affect anticoagulation control. Therefore, our null findings are all the more surprising; furthermore, they suggest the need to rethink quality improvement guidelines in anticoagulation care. For example, while some VA sites seem to emphasize face-to-face visits, this may not always be an efficient use of resources for the patient or for the provider, as telephone-based or mail-based care seemed to achieve similar results in this and other studies.[26,27]

Our study results do not suggest, nor do we believe, that efforts to improve the quality of oral anticoagulation therapy are futile. In fact, our group plans to implement a program to improve anticoagulation control in the VA, as we believe that this is feasible and important. Rather, our study findings suggest that we must look beyond the limited measures examined

herein to find the true determinants of high-quality oral anticoagulation care. Which aspects of structure of care did this study not measure and what might be fitting targets for future studies? We did not directly investigate how warfarin dosages are managed or which dosing protocols or computerized dosing aids are used. A rich literature shows that computer-aided dosing of warfarin[28-33] or simply judicious dosing in the absence of computerized support[34] can improve anticoagulation control. Our study was adequately powered to detect a 3% difference in RA-TTR, which is modest. By comparison, using a computerized dosing algorithm improves TTR by approximately 10% over usual care,[30] and more judicious dosing without computer assistance could improve TTR by approximately 6%.[34] It seems likely that aspects of structure or process of care that relate to actual warfarin dosage management would help explain the wide variations in ACC performance within the VA. Also, more detailed assessments of ACCs through site visits and staff interviews might identify differences between high-performing and low-performing clinics.[35]

This study has considerable strengths. We used measures of structure of care that are supported by prominent guidelines[1,7] and evaluated structure within the ACC itself rather than in the organizational culture of the entire medical center. In addition, our outcome measure (RA-TTR) has been carefully developed[8,14] and represents the state of the art in measuring intermediate outcomes of anticoagulation care.[2] This study also has limitations. Responses to the survey were by self-report, and it is possible that some responses (eg, the number of providers) were not fully accurate. Also, as already discussed, this survey instrument may not have collected enough detailed information to adequately characterize some aspects of structure, such as the nature rather than the mere existence of a quality improvement program. Finally, while our risk-adjustment model for TTR controlled for multiple measures of comorbidity and achieved a high R^2, it did not include data on adherence. However, part of the influence of good management may be expressed through improved adherence; adjusting for adherence could adjust away such an effect. Therefore, adherence arguably does not belong in the risk-adjustment model.

In summary, our data suggest that high-quality anticoagulation care can be provided across a wide array of structures of care. A dedicated ACC is a necessary first step for improving the quality of oral anticoagulation care.[3-6] However, within the structure of an ACC, further quality improvement efforts should focus on aspects of care that have been shown to affect anticoagulation control, such as the judicious dosing of warfarin.[34] A program of quality improvement based solely on the measures studied herein may not have the desired effect on outcomes.

Acknowledgment

The authors thank Fran Cunningham, PharmD, of VA Pharmacy Benefit Management for providing the data from the 2007 survey of VA Anticoagulation Clinics.

Author Affiliations: From the Center for Health Quality, Outcomes, and Economic Research (AJR, AO, JIR, DRB), Bedford VA Medical Center, Bedford, MA; Department of Medicine (AJR, EMH, ASA, DRB), Section of General Internal Medicine, Boston University School of Medicine, Boston, MA; Pharmacy Benefits Management (FC), VA Central Office, Washington, DC; Biostatistics Core (AO), Clinical Research Program, Children's Hospital Boston, Boston, MA; Department of Quantitative Health Sciences (ASA), Division of Biostatistics and Health Services Research, University of Massachusetts School of Medicine, Worcester, MA; Manchester VA Medical Center (PC), Manchester, NH; Providence VA Medical Center (MG), Providence, RI; and Department of Health Policy and Management (DRB), Boston University School of Public Health, Boston, MA.

Funding Source: This project was supported by a Career Development Award from the Veterans Affairs Health Services Research and Development Service. The opinions expressed herein do not necessarily represent the official views of the Department of Veterans Affairs.

Author Disclosures: Dr Hylek reports receiving honoraria from Bayer and Bristol-Myers Squibb and serving on advisory boards for Boehringer-Ingelheim, Bristol-Myers Squibb, Merck and Co, and sanofi aventis. The other authors (AJR, EMH, AO, ASA, JIR, PPC, MMG, DRB) report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (AJR, ASA, PPC, MMG, DRB); acquisition of data (AJR); analysis and interpretation of data (AJR, EMH, AO, ASA, JIR, PPC, MMG, DRB); drafting of the manuscript (AJR, AO, MMG, DRB); critical revision of the manuscript for important intellectual content (AJR, EMH, AO, ASA, JIR, PPC, DRB); statistical analysis (AJR, AO, ASA, JIR); obtaining funding (AJR); administrative, technical, or logistic support (AJR); and supervision (AJR, EMH).

Address correspondence to: Adam J. Rose, MD, MSc, Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, 200 Springs Rd, Bldg 70, Bedford, MA 01730. E-mail: adamrose@bu.edu.

REFERENCES

1. Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G; American College of Chest Physicians. Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th ed). *Chest.* 2008;133 (6)(suppl):160S-198S.

2. Rose AJ, Berlowitz DR, Frayne SM, Hylek EM. Measuring quality of oral anticoagulation care: extending quality measurement to a new field. *Jt Comm J Qual Patient Saf.* 2009;35(3):146-155.

3. Chiquette E, Amato MG, Bussey HI. Comparison of an anticoagulation clinic with usual medical care: anticoagulation control, patient outcomes, and health care costs. *Arch Intern Med.* 1998;158(15):1641-1647.

4. Cortelazzo S, Finazzi G, Viero P, et al. Thrombotic and hemorrhagic complications in patients with mechanical heart valve prosthesis attending an anticoagulation clinic. *Thromb Haemost.* 1993;69(4): 316-320.

5. Wilson SJ, Wells PS, Kovacs MJ, et al. Comparing the quality of oral anticoagulant management by anticoagulation clinics and by family physicians: a randomized controlled trial [published correction appears in *CMAJ.* 2004;170(4):451]. *CMAJ.* 2003;169(4):293-298.

6. Witt DM, Sadler MA, Shanahan RL, Mazzoli G, Tillman DJ. Effect of a centralized clinical pharmacy anticoagulation service on the outcomes of anticoagulation therapy. *Chest.* 2005;127(5):1515-1522.

7. Ansell JE, Oertel LB, Wittkowsky AK. *Managing Oral Anticoagulation Therapy: Clinical and Operational Guidelines.* 3rd ed. Conshohocken, PA: Wolters Kluwer Health; 2009.

8. Rose AJ, Hylek EM, Ozonoff A, Ash AS, Reisman JI, Berlowitz DR. Risk-adjusted percent time in therapeutic range as a quality indicator for outpatient oral anticoagulation: results of the Veterans Affairs Study to Improve Anticoagulation (VARIA). *Circ Cardiovasc Qual Outcomes.* 2011;4(1):22-29.

9. Connolly SJ, Pogue J, Eikelboom J, et al; ACTIVE W Investigators. Benefit of oral anticoagulant over antiplatelet therapy in atrial

fibrillation depends on the quality of international normalized ratio control achieved by centers and countries as measured by time in therapeutic range. *Circulation*. 2008;118(20):2029-2037.

10. van Leeuwen Y, Rosendaal FR, Cannegieter SC. Prediction of hemorrhagic and thrombotic events in patients with mechanical heart valve prostheses treated with oral anticoagulants. *J Thromb Haemost*. 2008;6(3):451-456.

11. van Walraven C, Oake N, Wells PS, Forster AJ. Burden of potentially avoidable anticoagulant-associated hemorrhagic and thromboembolic events in the elderly. *Chest*. 2007;131(5):1508-1515.

12. Veeger NJ, Piersma-Wichers M, Tijssen JG, Hillege HL, van der Meer J. Individual time within target range in patients treated with vitamin K antagonists: main determinant of quality of anticoagulation and predictor of clinical outcome: a retrospective study of 2300 consecutive patients with venous thromboembolism. *Br J Haematol*. 2005;128(4):513-519.

13. White HD, Gruber M, Feyzi J, et al. Comparison of outcomes among patients randomized to warfarin therapy according to anticoagulant control: results from SPORTIF III and V. *Arch Intern Med*. 2007;167(3):239-245.

14. Rose AJ, Hylek EM, Ozonoff A, Ash AS, Reisman JI, Berlowitz DR. Patient characteristics associated with oral anticoagulation control: results of the Veterans Affairs Study to Improve Anticoagulation (VARIA). *J Thromb Haemost*. 2010;8(10):2182-2191.

15. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briët E. A method to determine the optimal intensity of oral anticoagulant therapy. *Thromb Haemost*. 1993;69(3):236-239.

16. Department of Veterans Affairs. *Anticoagulation Therapy Management*. Washington, DC: Veterans Health Administration; February 2, 2009. VHA directive 2009-003.

17. Donabedian A. *Explorations in Quality Assessment and Monitoring*. Ann Arbor, MI: Health Administration Press; 1980.

18. Charbonneau A, Parker V, Meterko M, et al. The relationship of system-level quality improvement with quality of depression care. *Am J Manag Care*. 2004;10(11, pt 2):846-851.

19. Goldzweig CL, Parkerton PH, Washington DL, Lanto AB, Yano EM. Primary care practice and facility quality orientation: influence on breast and cervical cancer screening rates. *Am J Manag Care*. 2004; 10(4):265-272.

20. Hudson SV, Ohman-Strickland P, Cunningham R, Ferrante JM, Hahn K, Crabtree BF. The effects of teamwork and system support on colorectal cancer screening in primary care practices. *Cancer Detect Prev*. 2007;31(5):417-423.

21. Vaughn TE, McCoy KD, BootsMiller BJ, et al. Organizational predictors of adherence to ambulatory care screening guidelines. *Med Care*. 2002;40(12):1172-1185.

22. Yano EM, Soban LM, Parkerton PH, Etzioni DA. Primary care practice organization influences colorectal cancer screening performance. *Health Serv Res*. 2007;42(3, pt 1):1130-1149.

23. Subramanian U, Sutherland J, McCoy KD, Welke KF, Vaughn TE, Doebbeling BN. Facility-level factors influencing chronic heart failure care process performance in a national integrated health delivery system. *Med Care*. 2007;45(1):28-45.

24. Ward MM, Yankey JW, Vaughn TE, et al. Provider adherence to COPD guidelines: relationship to organizational factors. *J Eval Clin Pract*. 2005;11(4):379-387.

25. Ward MM, Yankey JW, Vaughn TE, et al. Physician process and patient outcome measures for diabetes care: relationships to organizational characteristics. *Med Care*. 2004;42(9):840-850.

26. Staresinic AG, Sorkness CA, Goodman BM, Pigarelli DW. Comparison of outcomes using 2 delivery models of anticoagulation care. *Arch Intern Med*. 2006;166(9):997-1002.

27. Wittkowsky AK, Nutescu EA, Blackburn J, et al. Outcomes of oral anticoagulant therapy managed by telephone vs in-office visits in an anticoagulation clinic setting. *Chest*. 2006;130(5):1385-1389.

28. Ageno W, Johnson J, Nowacki B, Turpie AG. A computer generated induction system for hospitalized patients starting on oral anticoagulant therapy. *Thromb Haemost*. 2000;83(6):849-852.

29. Manotti C, Moia M, Palareti G, Pengo V, Ria L, Dettori AG. Effect of computer-aided management on the quality of treatment in anticoagulated patients: a prospective, randomized, multicenter trial of APROAT (Automated Program for Oral Anticoagulant Treatment). *Haematologica*. 2001;86(10):1060-1070.

30. Poller L, Shiach CR, MacCallum PK, et al. Multicentre randomised study of computerised anticoagulant dosage: European Concerted Action on Anticoagulation. *Lancet*. 1998;352(9139):1505-1509.

31. Poller L, Wright D, Rowlands M. Prospective comparative study of computer programs used for management of warfarin. *J Clin Pathol*. 1993;46(4):299-303.

32. Ryan PJ, Gilbert M, Rose PE. Computer control of anticoagulant dose for therapeutic management. *BMJ*. 1989;299(6709):1207-1209.

33. Wilson R, James AH. Computer assisted management of warfarin treatment. *Br Med J (Clin Res Ed)*. 1984;289(6442):422-424.

34. Rose AJ, Ozonoff A, Berlowitz DR, Henault LE, Hylek EM. Warfarin dose management affects INR control. *J Thromb Haemost*. 2009;7(1): 94-101.

35. Utarini A, Winkvist A, Pelto GH. Appraising studies in health using rapid assessment procedures (RAP): eleven critical criteria. *Hum Organ*. 2001;60(4):390-400. ■

CHEST

Original Research

ANTITHROMBOTIC THERAPY

Reexamining the Recommended Follow-up Interval After Obtaining an In-Range International Normalized Ratio Value

Results from the Veterans Affairs Study to Improve Anticoagulation

Adam J. Rose, MD; Al Ozonoff, PhD; Dan R. Berlowitz, MD, MPH; Arlene S. Ash, PhD; Joel I. Reisman, AB; and Elaine M. Hylek, MD, MPH

Background: **Patients receiving oral anticoagulation therapy should be tested often enough to optimize control, but excessive testing increases burden and cost. We examined the relationship between follow-up intervals after obtaining an in-range (2.0-3.0) international normalized ratio (INR) and anticoagulation control.**

Methods: **We studied 104,451 patients who were receiving anticoagulation therapy from 100 anticoagulation clinics in the US Veterans Health Administration. Most patients (98,877) had at least one in-range INR followed by another INR within 56 days. For each such patient, we selected the last in-range INR and characterized the interval between this index value and the next INR. The independent variable was the site mean follow-up interval after obtaining an in-range INR. The dependent variable was the site mean risk-adjusted percentage of time in the therapeutic range (TTR).**

Results: **The site mean follow-up interval varied from 25 to 38 days. As the site mean follow-up interval became longer, the risk-adjusted TTR was worse (−0.51% per day, $P = .004$). This relationship persisted when the index value was the first consecutive in-range INR (−0.63%, $P < .001$) or the second (−0.58%, $P < .001$), but not the third or greater (−0.12%, $P = .46$).**

Conclusions: **Sites varied widely regarding follow-up intervals after obtaining an in-range INR (25-38 days). Shorter intervals were generally associated with better anticoagulation control, but after obtaining a third consecutive in-range value, this relationship was greatly attenuated and no longer statistically significant. Our results suggest that a maximum interval of 28 days after obtaining the first or second in-range value and consideration of a longer interval after obtaining the third or greater consecutive in-range value may be appropriate.**

CHEST 2011; 140(2):359–365

Abbreviations: E = expected percentage of time in the therapeutic range; INR = international normalized ratio; IQR = interquartile range; O = observed percentage of time in the therapeutic range; OAT = oral anticoagulation therapy; TTR = percentage of time in the therapeutic range; VA = Veterans Health Administration

Oral anticoagulation therapy (OAT) is used to treat or prevent thromboembolic disease for millions of patients.[1-6] Although OAT has been used for decades,[7] many aspects of this therapy are governed by tradition rather than evidence, including the maximum allowable follow-up interval. It is important to follow patients often enough to achieve good control, because better control is associated with improved outcomes.[8-12] However, we should not follow patients more often than necessary, because extra testing is costly and burdensome for patients and the medical system. In the absence of evidence, guideline committees have recommended different maximum follow-up intervals. The American College of Chest Physicians recommends a maximum interval of 28 days for all patients.[1] The American College of Cardiology/American Heart Association/European Society of Cardiology guideline

For editorial comment see page 281

allows up to 42 days when good control has been achieved for the patient.[2] The British Society of Haematology allows up to 90 days when good control has been achieved.[13] Underlying these divergent recommendations is a disagreement about whether frequent testing will capture important changes in the international normalized ratio (INR) or merely capture random noise and allow more possibilities for unwise and excessive dose adjustments.[14]

If more frequent testing were found to be unnecessary for a subset of patients with stable control, then increasing the follow-up interval for these patients would save money and reduce patient burden. We, therefore, set out to examine the impact of follow-up intervals after obtaining an in-range INR value (2.0-3.0) upon anticoagulation control for patients. Because patient-level follow-up intervals can be endogenously related to anticoagulation control (that is, these variables can exert a mutual influence on each other),[15,16] we studied follow-up intervals at the level of the site of care. We will show that this approach successfully addressed the problem of endogeneity. Our objective was to find empirical support for one or more of the existing guideline recommendations for the maximum follow-up interval.

Materials and Methods

Data

The database for this study has been described in detail.[17,18] The Veterans Affairs Study to Improve Anticoagulation (VARIA) included all patients who were deemed to be receiving OAT from the Veterans Health Administration (VA) between October 1, 2006, and September 30, 2008, according to the criteria described in the next sections. The study was approved by the institutional review board of the Bedford VA Medical Center.

Manuscript received October 23, 2010; revision accepted January 14, 2011.

Affiliations: From the Center for Health Quality, Outcomes, and Economic Research (Drs Rose, Ozonoff, and Berlowitz and Mr Reisman), Bedford VA Medical Center, Bedford; the Department of Medicine, Section of General Internal Medicine (Drs Rose, Berlowitz, Ash, and Hylek), and the Department of Health Policy and Management (Dr Berlowitz), Boston University School of Public Health; the Biostatistics Section (Dr Ozonoff), Boston Children's Hospital, Boston; and the Department of Quantitative Health Sciences, Division of Biostatistics and Health Services Research (Dr Ash), University of Massachusetts School of Medicine, Worcester, MA.

Funding/Support: This study was funded by a Career Development Award [CDA2-08-017] to Dr Rose from the Health Services Research and Development Service of the US Department of Veterans Affairs.

Correspondence to: Adam J. Rose, MD, Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, 200 Springs Rd, Bldg 70, Bedford, MA 01730; e-mail: adamrose@bu.edu

DOI: 10.1378/chest.10-2738

Sites of Care

We studied 100 VA sites of care, each of which has a specialized anticoagulation clinic run by clinical pharmacists under the supervision of a medical director.[19] For this study, we excluded INR values from the first 6 months of therapy with warfarin (the inception period). We have previously shown that the percentage of time in the therapeutic range (TTR) is lower during the inception period,[16-18] and follow-up intervals are justifiably shorter during this period.

Dependent Variable: Site Risk-Adjusted TTR

Our dependent variable was mean site risk-adjusted for the TTR, a measure of anticoagulation control over time. We calculated the TTR using the Rosendaal method, which uses linear interpolation to assign an INR value to each day between successive observed INR values.[20] Gaps of >56 days between INR values are not interpolated. After interpolation, the percentage of time during which the interpolated INR values are between 2.0 and 3.0 is calculated.[20] We calculated the risk-adjusted TTR for all patients, including the small subset that did not record an in-range INR value.

We included INR values obtained when patients were given warfarin, that is, when a patient was either in possession of warfarin or having INR tests at least every 42 days. We defined the period of warfarin possession as the duration of the most recent VA prescription for warfarin plus 30 days. We excluded INR tests measured while the patient was hospitalized within the VA system. We also excluded patients whose primary indication to receive warfarin was valvular heart disease or the use of a prosthetic heart valve. Many such patients have a target INR range of 2.5 to 3.5 rather than the more standard 2.0 to 3.0, but it is not possible to determine with certainty which patients have the higher target range. Without specific knowledge of the target range, we cannot calculate the TTR.

Our risk adjustment model for the TTR, which controls for patient-level risk factors for poor TTR, has been described previously.[17,18] Site risk-adjusted TTR was calculated using the following procedure. First, for each patient, we calculated the observed TTR (O) and applied the risk-adjustment model to calculate the expected TTR (E). Then, an observed minus expected (O − E) score was calculated for each patient. The mean O − E score for each site constituted its risk-adjusted TTR.

Independent Variables: Site Mean Follow-up Intervals

We located all patients who had at least one in-range INR value (2.0-3.0) followed by a second INR value 14 to 56 days later, without an intervening hospitalization. We judged that obtaining a follow-up INR value within <2 weeks was unlikely to represent usual treatment and was more likely related to an event such as an emergency visit or an outpatient procedure. We excluded intervals >56 days because they could not be interpolated for the calculation of the TTR.[20] To examine the impact of these decisions, we also reran our main analyses using different qualifying intervals, including 1 to 56 days, 14 to 90 days, and 1 to 90 days; our main results did not change.

When a patient had multiple qualifying episodes, we arbitrarily selected the last such episode, so that each patient was sampled no more than once. We examined the number of days between the index INR (ie, the in-range value) and the following INR. We averaged values from individual patients to calculate a mean value for each site. We also looked backward from each index INR value to see whether the INR immediately before it was also in range, and, if so, also looked at the INR before that. Thus, we divided patients into three mutually exclusive groups based on whether

the index INR value was the first, second, or third consecutive in-range value. We stratified the sample on these three categories and calculated the mean follow-up interval within each stratum. Thus, for each site, we calculated the mean follow-up interval after obtaining a single in-range INR value, two consecutive INR values, and three consecutive INR values.

Statistical Analyses

We examined the baseline characteristics of the patients in our source population (all patients who received OAT) and our study sample (patients with at least one in-range value). We examined the relationship between follow-up intervals and risk-adjusted TTR at the level of the individual patient. We then examined this relationship at the site level, using both simple correlation and linear regression. Finally, we examined the differential effect on anticoagulation control of the mean site follow-up interval after obtaining one, two, or three consecutive in-range INR values. All analyses were conducted using SAS, version 9.1 (SAS Institute Inc; Cary, North Carolina).

Results

Patient Population and Anticoagulation Control

Baseline characteristics for the study patients are described in Table 1, and they were similar between the source population (104,451 patients who received OAT) and the study sample (98,877 patients with at least one in-range INR). The sample patients were mostly men (98%) and had a median age of 73 years. Patients had a substantial burden of comorbidity. For example, 40% had diabetes, 32% had heart failure, and 14% had chronic kidney disease. The burden of mental illness and substance abuse was considerable: 21% had major depression, 9% abused alcohol, and 5% had dementia. The mean TTR for the study sample was 63%.

There were 100 sites of care. The site mean TTR (O) ranged from 41% to 72%. The site expected TTR (E) ranged from 58% to 65%. The site risk-adjusted TTR (O − E) score ranged from 19% below to 12% above expected.

Follow-up Intervals

We characterized the follow-up intervals after stratifying by the index INR (ie, the INR value that preceded the interval). As would be expected, the follow-up intervals were longest when the index value was in range (Fig 1) and were uniform throughout that range. This confirmed our decision to treat all INR values within the target range (2.0-3.0) equally.

Patient-Level Analysis

The mean interval after obtaining an in-range INR was 29.8 days, and the mean TTR was 63%. For each additional day of the follow-up interval, the patient experienced a 0.35% higher TTR ($P < .001$) and a 0.25% higher TTR after adjustment for covariates ($P < .001$). The effect was therefore in the expected direction (ie, longer follow-up intervals were associated with improved control on the patient level).

Table 1—*Baseline Sample Characteristics for Source Population and Study Sample*

Characteristic	Source Population (N = 104,451)	Study Sample (n = 98,877)
Female sex	1.9	1.9
Median age (IQR)	72 (62-79)	73 (63-79)
Race/ethnicity		
Non-Hispanic white	77.2	77.5
Non-Hispanic black	8.5	8.1
Hispanic	2.8	2.8
Asian	0.3	0.3
Native American	0.3	0.3
Other/unknown	10.9	11.0
Median percentage of poverty in zip code of residence (IQR)	10.7 (6.6-15.9)	10.6 (6.5-15.8)
Median distance from nearest VA facility, miles (IQR)	7.8 (3.7-16.5)	7.8 (3.7-16.4)
Primary indication for warfarin[a]		
Atrial fibrillation	64.2	64.7
VTE	27.3	26.9
All others combined	8.5	8.4
Physical comorbid conditions		
Newly diagnosed cancer	6.8	6.6
Chronic kidney disease	14.2	13.9
Chronic liver disease	1.2	1.1
Chronic lung disease	29.4	29.0
Diabetes	40.1	39.9
Epilepsy	2.8	2.7
Heart failure	32.8	32.4
Hyperlipidemia	75.4	75.7
Hypertension	84.0	84.1
Mental comorbid conditions		
Alcohol abuse	9.3	8.9
Bipolar disorder	2.3	2.2
Dementia	5.3	5.2
Major depression	21.6	21.2
Nonalcohol substance abuse	4.0	3.7
Median number of medications (IQR)	8 (6-12)	8 (6-12)
Hospitalized at least once	26.2	25.3
Anticoagulation control		
TTR, mean (SD)	61.2 (21.9)	63.4 (19.8)

Data on the source population were used to calculate site performance (ie, risk-adjusted TTR), and data on the study sample were used to characterize the follow-up interval after obtaining an in-range (2.0-3.0) INR value. Data are presented as % unless otherwise indicated. INR = international normalized ratio; IQR = interquartile range; TTR = percentage of time in the therapeutic range; VA = Veterans Health Administration.

[a]Patients whose main indication for anticoagulation was valvular heart disease or the use of a prosthetic heart valve were excluded from this study.

Site-Level Analysis

There were 100 sites, with a median of 888 patients per site (interquartile range [IQR], 576-1,341). The mean site follow-up interval after obtaining an in-range

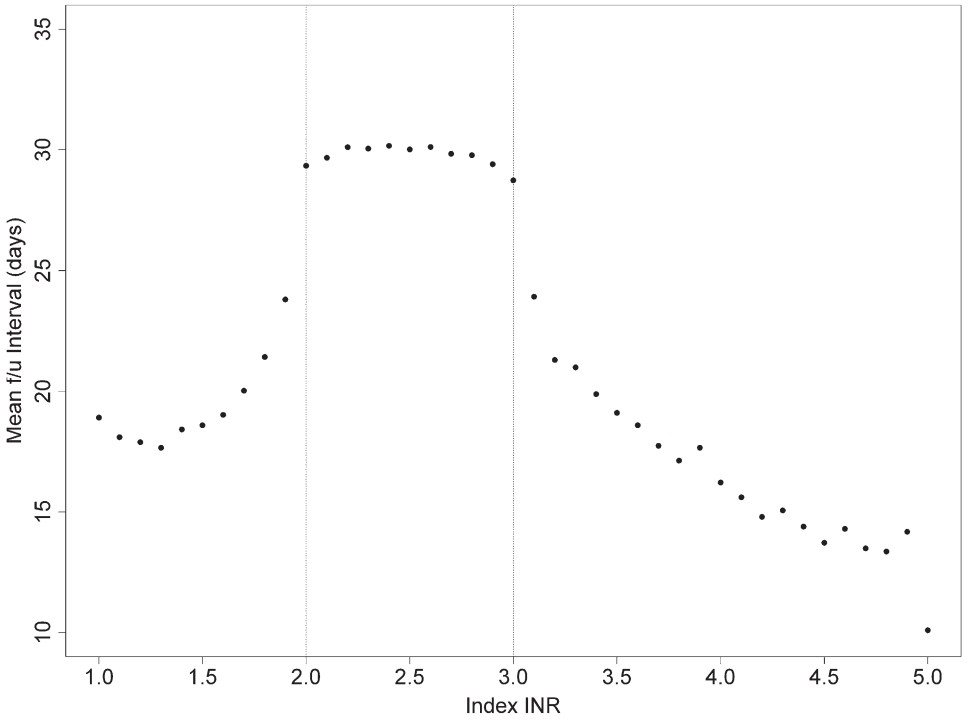


FIGURE 1. The interval (in days) between INR tests stratified by the index INR value (N = 104,451). Each patient contributes one interval (the patient's last in the database). All INR values ≥5.0 are combined into a single data point. f/u = follow-up; INR = international normalized ratio.

INR varied from 25 to 38 days (Fig 2). A longer mean follow-up interval after obtaining any in-range INR value was associated with worse performance, the opposite effect of that seen at the patient level (Fig 2; r = −0.28, $P = .004$). For each day of follow-up interval after obtaining an in-range INR, the site O−E score was 0.51% lower ($P = .004$) (Table 2).

We subdivided the index INR values into the first, second, or third consecutive in-range values. We found that 47% of the index values were preceded by an out-of-range value, while 23% were preceded by a single in-range value, and 31% were preceded by at least two consecutive in-range values (Table 2). The relationship between longer intervals and poorer site-level performance held for the interval after obtaining the first consecutive in-range value (−0.63%, $P < .001$) and the second consecutive in-range value (−0.58%, $P < .001$), but was greatly attenuated and no longer statistically significant after obtaining the third or greater consecutive in-range value (−0.12%, $P = .46$). Similar results were seen among patients aged >75 years and patients who received anticoagulation treatment for atrial fibrillation.

Discussion

In our study, 100 sites in an integrated health system varied considerably regarding the follow-up interval after obtaining an in-range (2.0-3.0) INR value. This variation likely reflects a lack of evidence and divergent practice guidelines.[1,2,13] We took advantage of these variations in practice to examine the relationship between follow-up intervals and anticoagulation control. We found that longer follow-up intervals predicted better anticoagulation control at the level of the individual patient but worse anticoagulation control at the level of site of care. This apparent paradox, which is what we expected to find, can be explained as follows.

Each site of care is managed by a group of clinicians who, whether by official policy or unofficial tradition, have arrived at a usual follow-up interval after obtaining an in-range INR. However, this usual interval may be lengthened or shortened depending on the clinician's suspicion that a particular patient will be in range at the next visit. Because these clinical impressions have some validity, longer follow-up intervals at the patient level predict better anticoagulation control.[16] However, by examining the mean follow-up intervals at the site level, we showed that when the site had a pattern of longer follow-up intervals, the intervals were actually associated with worse anticoagulation control.

In a previous study,[18] we showed that patients who receive anticoagulation treatment at VA facilities have a mean TTR of 58%; our goal is to improve this to approximately 70%. An improvement of this magnitude would be expected to improve patient outcomes

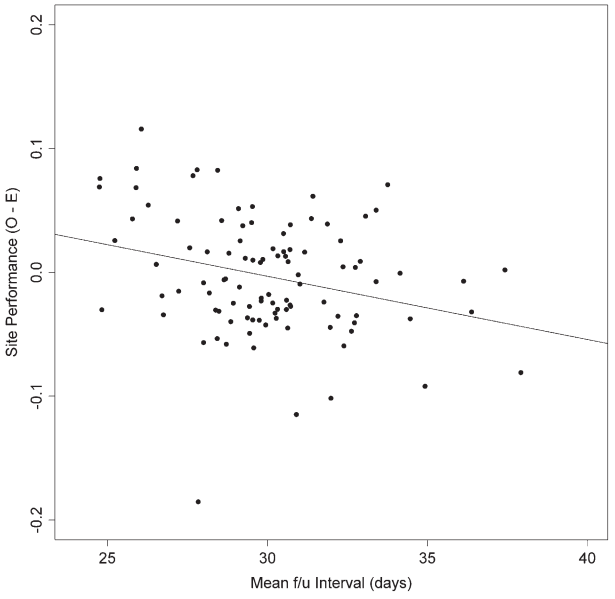


FIGURE 2. The correlation between the mean site-level follow-up interval after obtaining an in-range INR value (2.0-3.0) and the site-level performance as measured by the risk-adjusted percentage of time in the therapeutic range (TTR) (No. = 100 sites). For the regression line, r = −0.28 (P = .004), indicating that shorter follow-up intervals are associated with better site-level performance. O − E = observed TTR minus expected TTR. See Figure 1 legend for expansion of the other abbreviations.

considerably.[8-12] In the present study, the site mean follow-up intervals varied from 25 to 38 days. Because each additional day in the follow-up interval reduces the site-level TTR by 0.51%, this variation would be expected to account for a 6.6% difference in the TTR across the range of practice we observed. These results suggest that optimizing the follow-up intervals after obtaining in-range INR values could help achieve our goal of improving the TTR to 70%. Our main finding (that longer follow-up intervals were associated with poorer control) persisted when we limited our analysis to patients aged > 75 years and those who received anticoagulation treatment for atrial fibrillation. Given the particular importance of excellent control for improving outcomes in these groups,[2,5,8,12] they may benefit most from efforts to improve TTR by optimizing the follow-up interval.

In addition, we examined follow-up intervals after obtaining the first, second, or third or greater consecutive in-range INR value. The association between longer follow-up intervals and poorer performance was seen only after obtaining a first or second consecutive in-range value, but after obtaining a third or greater consecutive in-range value, it was greatly attenuated and no longer statistically significant. This provides an opportunity to explain our main findings in the context of divergent guideline recommendations. In general, this study lends support to the American College of Chest Physicians recommendation to limit follow-up intervals for many patients to 28 days,[1] because sites of care that pursued this strategy in our study had the best anticoagulation control (Fig 2). However, our results also suggest that for patients with extremely stable control, extending the follow-up interval beyond 28 days might be considered. Several previous studies have also supported the idea that some patients can safely extend their

Table 2—*Relationship Between Site Mean Follow-up Interval After Obtaining an In-Range INR (2.0-3.0) and Site Performance*

Index INR Value	Group, No. (%)	Mean Follow-up Interval, d (SD)	Effect on Site Performance, %	*P* Value
All patients				
All in-range values	98,877 (100)	29.8 (8.8)	−0.51	0.004
First consecutive in-range value	46,386 (47)	28.2 (9.2)	−0.63	< 0.001
Second consecutive in-range value	22,260 (23)	30.4 (8.3)	−0.58	< 0.001
Third or greater consecutive in-range value	30,231 (31)	31.9 (7.9)	−0.12	0.46
Patients aged > 75 y				
All in-range values	38,299 (100)	29.6 (8.5)	−0.48	0.007
First consecutive in-range value	17,356 (45)	28.0 (8.9)	−0.52	< 0.001
Second consecutive in-range value	8,598 (22)	30.1 (8.0)	−0.63	< 0.001
Third or greater consecutive in-range value	12,345 (32)	31.6 (7.6)	−0.21	0.18
Patients who received anticoagulation treatment for atrial fibrillation				
All in-range values	63,978 (100)	30.0 (8.6)	−0.53	0.003
First consecutive in-range value	29,353 (46)	28.4 (9.1)	−0.60	< 0.001
Second consecutive in-range value	14,454 (23)	30.5 (8.2)	−0.59	< 0.001
Third or greater consecutive in-range value	20,171 (32)	32.0 (7.8)	−0.19	0.24

Data were measured by risk-adjusted TTR. The index (in-range) INR value was stratified by whether it was the first, second, or third or greater consecutive in-range value. Analyses were performed first using the entire dataset, followed by selected subsets (ie, patients aged > 75 years, patients with atrial fibrillation). The effect on site performance was determined in units of TTR per additional day of follow-up interval. For example, an effect size of −1.0 signifies a 1% decrease in the site mean TTR per additional day of follow-up interval. See Table 1 legend for expansion of abbreviations.

follow-up interval beyond 28 days.[16,21] Most notably, Witt et al[21] showed that patients with all of their INR values in range had lower rates of complications than the comparator patients, despite the fact that they also had fewer INR tests. Indeed, an ongoing randomized trial is comparing a 1-month follow-up interval with a 3-month follow-up interval for patients whose warfarin dose has been stable for at least 6 months,[22] a group likely to overlap considerably with patients who record three consecutive in-range values. If a longer follow-up interval is confirmed to be safe for such patients, this could save time and money for patients and the health-care system.

This study has important strengths. The database comprised over 100,000 patients and was rich in clinical detail. Our outcome measure (risk-adjusted TTR) is the product of extensive development[17,18] and represents the state of the art in quality measurement in OAT. We used innovative methods to avoid the problem of endogeneity in observational studies of follow-up intervals. However, our study also has limitations. As with any observational study, inferences about cause and effect must be interpreted with caution. A second limitation is that we measured the follow-up interval that was achieved, not necessarily the interval that the clinician requested. It is possible that some patients were followed up later than the clinician wanted and that this negatively impacted anticoagulation control. We addressed this by risk adjusting the risk for patient characteristics that impact anticoagulation control,[17] thereby equalizing this effect between sites. Furthermore, our finding that longer follow-up intervals are associated with better control on the patient level would also argue against this being an issue. A third limitation is that this study was limited to patients with a target INR of 2.0 to 3.0 and at least 6 months of experience with warfarin; our results may not apply to patients with different target ranges or those new to warfarin. Finally, our sample population was predominantly male and had a high degree of comorbidity. However, it is unclear how this would alter the basic relationships we demonstrated between follow-up intervals and anticoagulation control.

In summary, we found that 100 sites of care within an integrated health system (the VA) pursued widely divergent follow-up intervals after obtaining an in-range INR value. Longer follow-up intervals are associated with poorer anticoagulation control at the site level, except when three or more consecutive in-range values have been recorded for the patient. Our results support a 28-day maximum follow-up interval after obtaining one or two consecutive in-range values and consideration of a longer interval after obtaining a third or greater in-range value. Our study also suggests that reducing practice variation could contribute both to improved anticoagulation control and enhanced efficiency.

Acknowledgments

Author contributions: Dr. Rose had full access to the data and will vouch for the integrity of the data analysis.
Dr Rose: contributed to the development of the study idea, obtaining of funds, data collection, analysis and interpretation of the data, drafting of the manuscript, revision of the manuscript for important intellectual content, and study supervision.
Dr Ozonoff: contributed to the statistical supervision of the data, analysis and interpretation of the data, and revision of the manuscript for important intellectual content.
Dr Berlowitz: contributed to the obtaining of funds, analysis and interpretation of the data, revision of the manuscript for important intellectual content, and study supervision.
Dr Ash: contributed to the analysis and interpretation of the data and revision of the manuscript for important intellectual content.
Mr Reisman: contributed to data collection, statistical programming, analysis and interpretation of the data, and revision of the manuscript for important intellectual content.
Dr Hylek: contributed to the analysis and interpretation of the data, revision of the manuscript for important intellectual content, and study supervision.
Financial/nonfinancial disclosures: The authors have reported to *CHEST* the following conflicts of interest: Dr Rose is supported by a Career Development Award from the Health Services Research and Development Service of the US Department of Veterans Affairs. Dr Hylek has received honoraria from Bayer and Bristol-Myers Squibb and has served on advisory boards for Boehringer-Ingelheim, Bristol-Myers Squibb, and Sanofi-Aventis. Drs Ozonoff, Berlowitz, and Ash and Mr Reisman have reported that no potential conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.
Role of sponsors: The US Department of Veterans Affairs had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; and the preparation, review, and approval of the manuscript. The opinions expressed in this manuscript do not necessarily represent the official views of the US Department of Veterans Affairs.

References

1. Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G; American College of Chest Physicians. Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians evidence-based clinical practice guidelines (8th edition). *Chest*. 2008;133(suppl 6):160S-198S.
2. Fuster V, Rydén LE, Cannom DS, et al; American College of Cardiology/American Heart Association Task Force on Practice Guidelines; European Society of Cardiology Committee for Practice Guidelines; European Heart Rhythm Association; Heart Rhythm Society. ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Revise the 2001 Guidelines for the Management of Patients With Atrial Fibrillation): developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. *Circulation*. 2006; 114(7):e257-e354.
3. Kearon C, Kahn SR, Agnelli G, Goldhaber S, Raskob GE, Comerota AJ; American College of Chest Physicians. Antithrombotic therapy for venous thromboembolic disease: American College of Chest Physicians evidence-based clinical practice guidelines (8th edition). *Chest*. 2008;133(suppl 6): 454S-545S.

4. Salem DN, O'Gara PT, Madias C, Pauker SG; American College of Chest Physicians. Valvular and structural heart disease: American College of Chest Physicians evidence-based clinical practice guidelines (8th edition). *Chest*. 2008; 133(suppl 6):593S-629S.
5. Singer DE, Albers GW, Dalen JE, et al; American College of Chest Physicians. Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians evidence-based clinical practice guidelines (8th edition). *Chest*. 2008;133(suppl 6): 546S-592S.
6. Wysowski DK, Nourjah P, Swartz L. Bleeding complications with warfarin use: a prevalent adverse effect resulting in regulatory action. *Arch Intern Med*. 2007;167(13):1414-1419.
7. Hirsh J, Bates SM. Clinical trials that have influenced the treatment of venous thromboembolism: a historical perspective. *Ann Intern Med*. 2001;134(5):409-417.
8. Connolly SJ, Pogue J, Eikelboom J, et al; ACTIVE W Investigators. Benefit of oral anticoagulant over antiplatelet therapy in atrial fibrillation depends on the quality of international normalized ratio control achieved by centers and countries as measured by time in therapeutic range. *Circulation*. 2008; 118(20):2029-2037.
9. van Leeuwen Y, Rosendaal FR, Cannegieter SC. Prediction of hemorrhagic and thrombotic events in patients with mechanical heart valve prostheses treated with oral anticoagulants. *J Thromb Haemost*. 2008;6(3):451-456.
10. van Walraven C, Oake N, Wells PS, Forster AJ. Burden of potentially avoidable anticoagulant-associated hemorrhagic and thromboembolic events in the elderly. *Chest*. 2007;131(5): 1508-1515.
11. Veeger NJ, Piersma-Wichers M, Tijssen JG, Hillege HL, van der Meer J. Individual time within target range in patients treated with vitamin K antagonists: main determinant of quality of anticoagulation and predictor of clinical outcome. A retrospective study of 2300 consecutive patients with venous thromboembolism. *Br J Haematol*. 2005;128(4): 513-519.
12. White HD, Gruber M, Feyzi J, et al. Comparison of outcomes among patients randomized to warfarin therapy according to anticoagulant control: results from SPORTIF III and V. *Arch Intern Med*. 2007;167(3):239-245.
13. Baglin TP, Keeling DM, Watson HG; British Committee for Standards in Haematology. Guidelines on oral anticoagulation (warfarin): third edition-2005 update. *Br J Haematol*. 2006; 132(3):277-285.
14. Rose AJ, Ozonoff A, Berlowitz DR, Henault LE, Hylek EM. Warfarin dose management affects INR control. *J Thromb Haemost*. 2009;7(1):94-101.
15. Fitzmaurice DA. Oral anticoagulation should be managed in the community with treatment aimed at standard therapeutic targets and increased recall intervals. *J Thromb Haemost*. 2008;6(10):1645-1646.
16. Rose AJ, Ozonoff A, Henault LE, Hylek EM. Warfarin for atrial fibrillation in community-based practise. *J Thromb Haemost*. 2008;6(10):1647-1654.
17. Rose AJ, Hylek EM, Ozonoff A, Ash AS, Reisman JI, Berlowitz DR. Patient characteristics associated with oral anticoagulation control: results of the Veterans AffaiRs Study to Improve Anticoagulation (VARIA). *J Thromb Haemost*. 2010;8(10):2182-2191.
18. Rose AJ, Hylek EM, Ozonoff A, Ash AS, Reisman JI, Berlowitz DR. Risk-adjusted percent time in therapeutic range as a quality indicator for outpatient oral anticoagulation: results of the Veterans AffaiRs study to Improve Anticoagulation (VARIA). *Circ Cardiovasc Qual Outcomes*. 2011;4(1): 22-29
19. Department of Veterans Affairs, Veterans Health Administration. Anticoagulation therapy management. Washington, DC: US Department of Veterans Affairs; 2009. VHA Directive 2009-003.
20. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briët E. A method to determine the optimal intensity of oral anticoagulant therapy. *Thromb Haemost*. 1993;69(3):236-239.
21. Witt DM, Delate T, Clark NP, et al; Warfarin Associated Research Projects and Other Endeavors (WARPED) Consortium. Outcomes and predictors of very stable INR control during chronic anticoagulation therapy. *Blood*. 2009;114(5): 952-956.
22. National Institutes of Health Clinical Center. Prolongation of the interval between monitoring of warfarin in stable patients. NCT00356759. ClinicalTrials.gov. Bethesda, MD: National Institutes of Health; 2010. http://www.clinicaltrials.gov.

Reexamining the Recommended Follow-up Interval After Obtaining an In-Range International Normalized Ratio Value : Results from the Veterans Affairs Study to Improve Anticoagulation

Adam J. Rose, Al Ozonoff, Dan R. Berlowitz, Arlene S. Ash, Joel I. Reisman and Elaine M. Hylek

Chest 2011;140; 359-365; Prepublished online February 10, 2011;
DOI 10.1378/chest.10-2738

Journal of Health Communication, 16:9–10, 2011

ISSN: 1081-0730 print/1087-0415 online
DOI: 10.1080/10810730.2011.605824

Routledge
Taylor & Francis Group

Guest Editorial

Incorporating Health Literacy into Larger Operational Environments

STEVEN R. RUSH

United Health Group Health Literacy Innovations Program, Minnetonka, Minnesota, USA

MICHAEL K. PAASCHE-ORLOW

Section of General Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

In their article, *Health Literacy and Health Care Spending and Utilization in a Consumer-Driven Health Plan*, Hardie et al. examine health literacy and health care spending and utilization by linking responses to three health literacy screening questions to 2006 claims data of enrollees new to consumer-driven health plans. Focusing on emergency department and inpatient admission data, they were able to link costs for these services with self-reported ability to read and learn about medical conditions. They found that better health literacy was associated with fewer inpatient admissions, fewer emergency department visits, and lower total medical spending. Their data supports the notion that members with lower health literacy tended to access care with more advanced conditions and consequently required more services and more costly services. However, as they acknowledge, their analyses are limited by the absence of several parameters such as income. They also do not have data on the appropriateness of hospital or emergency department utilization. The opportunity to use existing data such as these is frequently compromised by missing data elements.

Several things about this project deserve comment. The article represents the collaboration between a health insurer and independent investigators and capitalizes on claims data. This is a welcome sign. Insurers and health plans have become increasingly aware of the need to provide accessible, understandable, and actionable communications to their consumers. Health plans, individually and in conjunction with America's Health Insurance Plans (AHIP), have begun working to incorporate the principles of clear health communication into their organizations (America's Health Insurance Plans, 2011). However, much progress is needed. Little has been done to transfer the findings of health literacy research into the larger operational

Address correspondence to Michael K. Paasche-Orlow, Department of Medicine, Boston University School of Medicine, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA. E-mail: Michael.PaascheOrlow@bmc.org

environment of health plans. Dissemination in these environments is driven by opportunity, leadership, policy, competition, and the bottom line. While the dominant TLA (three-letter acronym) in the research community is EBM (evidence-based medicine), in the health care industrial sector it is ROI (return-on-investment). The emerging Health Literacy research community can aspire to influence both. To do so will require the types of basic research called for by Johnson and colleagues in their Commentary in this issue, but it will also require research in cost effectiveness, organizational change, and implementation. As Allen et al. describe in their article in this issue, a lot of work is needed to establish the generalizability and sustainability of health literacy interventions. As much time as we spend striving to publish excellent articles we have to remember that the Journal Impact Factor is not the goal of research—it is the human impact factor that matters.

Second, Hardie et al. used the Chew questions to measure health literacy. By and large, subjective measures that have been validated against pyrite standard measures such as the REALM and TOFHLA represent a severe limitation and should be eschewed. After all, as Pleasant et al. write in their Commentary in this special issue, such measures are not based in any theory of health literacy, fail to measure all the domains of the underlying concepts, and have not been rigorously tested according to appropriate psychometric methods. And yet, these limited tools appear to be measuring something that is very important. In fact, as much as we agree with the motivation to significantly improve measurement in this field, the relevance of additional domains that would be included (e.g., navigation) will need to be empirically established. However crude, some research aims do not require more refined measures and many opportunities for valuable research and evaluation will be missed if we completely shun existing tools.

For the aim of showing that difficulty with forms is associated with higher costs, these screening questions were sufficient. But they do not provide adequate guidance for the subsequent question: What should be done? And even once a causal pathway is exposed and an intervention is designed and proven to be effective, further work is needed to transfer and maintain these interventions into larger operational settings. Greater collaboration between health literacy researchers and thought leaders and payers of health care is needed. Such collaboration could provide access to data, funding for research, and the levers to inform change within diverse parts of the health care system.

Consumers and their employers recognize the need for accessible, understandable, and actionable health and wellness communications. The paper by Hardie et al. helps show the business case; this will help influence insurers to get involved.

References

America's Health Insurance Plans (AHIP). (2011). Health Literacy: A Toolkit for Communicators. Retrieved July 12, 2011, from http://www.ahip.org/content/default.aspx?bc=39|341|22050

Hardie, N. A., Kyanko, K., Busch, S., LoSasso, A. T., & Levin, R. A. (2011). Health literacy and health care spending and utilization in a consumer-driven health plan. *Journal of Health Communication*, *16*(S3), 307–320.

Belief in AIDS Origin Conspiracy Theory and Willingness to Participate in Biomedical Research Studies: Findings in Whites, Blacks, and Hispanics in Seven Cities Across Two Surveys

Stefanie L. Russell,[1] Ralph V. Katz,[1] Min Qi Wang,[2] Ryan Lee,[3] B. Lee Green,[4] Nancy R. Kressin,[5,6] and Cristina Claudio[7]

[1]Department of Epidemiology and Health Promotion, New York University, College of Dentistry, New York, New York, USA; [2]Department of Public and Community Health, University of Maryland, College Park, Maryland, USA; [3]New York University, College of Dentistry, New York, New York, USA; [4]Office of Institutional Diversity, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida, USA; [5]Center for Health Quality, Outcomes & Economic Research, Department of Veterans Affairs, Bedford VA, Bedford, Massachusetts, USA; [6]Department of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts, USA; [7]Community Dentistry Section, School of Dentistry, University of Puerto Rico, Medical Sciences Campus, San Juan, Puerto Rico

Purpose: The purpose of this study was to determine whether a belief in the AIDS origin conspiracy theory is related to likelihood or fear of participation in research studies. **Methods:** The Tuskegee Legacy Project Questionnaire was administered via random-digit-dialed telephone interview to black, white, and Hispanic participants in 4 cities in 1999 and 2000 (n = 1,133) and in 3 cities in 2003 (n = 1,162). **Results:** In 1999, 27.8% of blacks, 23.6% of Hispanics, and 8% of whites ($P \leq .001$) reported that it was "very or somewhat likely" that AIDS is "the result of a government plan to intentionally kill a certain group of people by genocide." In 2003, 34.1% of blacks, 21.9% of Hispanics, and 8.4% of whites ($P \leq .001$) reported the same. **Conclusions:** Whereas blacks and Hispanics were more than 3 times more likely than whites to believe in this AIDS origin conspiracy theory, holding this belief was not associated with a decreased likelihood of participation in, or increased fear of participation in, biomedical research. **Key words:** *AIDS, biomedical research, conspiracy theory, health disparities, HIV, research participation*

Despite the 1994 US government legislation that mandated inclusion of racial and ethnic minorities in all nationally funded biomedical research studies, minority groups remain underrepresented.[1] Reasons for this disparity that have been previously examined with varying results include fear of participation in biomedical research,[2,3] mistrust in doctors/scientists, mistrust in the US government,[4] perceived racism,[5] lack of awareness and/or knowledge of the availability of clinical trials,[3,6] and knowledge of past research abuses, including the US Public Health Service (USPHS)-Tuskegee Syphilis Study.[2,7–12]

An additional, underexplored potential explanation for the lack of participation in biomedical research by racial and ethnic minorities in the United States may be a belief in conspiracy theories, specifically regarding the origin of AIDS and the role of government in the HIV epidemic, which have been shown to be held by a substantial proportion of persons in these minority groups.[13] Although AIDS conspiracy beliefs have been shown to be associated with health-damaging behaviors, including inconsistent condom usage and greater numbers of sexual partners,[14,15] and despite persistent reported racial/ethnic disparity in biomedical research participation[4,16–18] and the documented popularity of conspiracy theories regarding the origins of AIDS in minority populations,[13–15,19–30] to date only one study has examined

Address for correspondence: Stefanie L. Russell, DDS, MPH, PhD, Department of Epidemiology and Health Promotion, New York University, College of Dentistry, 250 Park Avenue South, Room 635, New York, NY 10003 USA. E-mail: stefanie.russell@nyu.edu

HIV Clin Trials 2011;12(1):37–47

www.thomasland.com

doi: 10.1310/hct1201-37

the relationship between belief in conspiracy theories of the origin of AIDS and willingness to participate in biomedical research studies.[31] In this study, the relationship of misconceptions about HIV and HIV vaccines to willingness to participate in HIV vaccine research trials was investigated among a group of 220 community college students in Atlanta, Georgia. Racial minority students were more likely than white students to believe that "the AIDS epidemic was caused by a government conspiracy." Belief in this statement was unrelated to willingness to participate in research. However, it was unclear whether results varied by race/ethnicity. Also, belief in AIDS origin conspiracy was not analyzed separately from other "AIDS misconception" questions, but as part of a scale. It is uncertain to what extent the results of this previous study might be applicable to the broader question of willingness to participate in biomedical research studies in general, or to other populations, given the narrow sampling framework (ie, the inclusion of only white, African American, and Asian college students) and the lack of demographic information provided about the participants of this study (eg, age, socioeconomic status).

The goals of the present analyses were as follows: (1) to examine and compare the proportion of blacks, whites, and Hispanic participants in 2 large, multicity population-based cohorts who believe that the origin of AIDS is a government conspiracy ("AIDS origin conspiracy theory"); (2) to determine which demographic factors within each group are related to belief in AIDS origin conspiracy theory; and (3) to evaluate whether belief in AIDS origin conspiracy theory is related to an expressed fear of participation and/or a decreased likelihood of participation in biomedical research within each racial/ethnic group. We hypothesized that blacks and Hispanics would be more likely to report a belief in AIDS origin conspiracy theory, compared with non-Hispanic whites, and that believing in AIDS origin conspiracy would be related to fear of participation and likelihood of participation in biomedical research studies, regardless of racial/ethnic identification.

METHODS

The data for this report were obtained by the repeated use of the Tuskegee Legacy Project (TLP) Questionnaire administered in 2 separate random-digit-dialed (RDD) telephone surveys conducted in a rigorous and similar fashion in 1999 and 2000 in 4 US cities and in 2003 in 3 US cities. The 60-item, 25-minute-long TLP survey was developed via pilot studies conducted from 1994 to 1997 by a multidisciplinary, multi-university research team, and its design, including the rationale for the selection of questions and cities, has been described previously.[2] Questions were initially drawn from other studies, and the survey was extensively piloted with members of the target population and refined over time.[2] In late 1999 and early 2000, the TLP Questionnaire was administered by the Survey Research Unit of the University of Alabama to targeted populations of noninstitutionalized self-reported blacks, whites, or Hispanics (of either race) aged 18 years or older living in households with working telephones in Birmingham, Alabama; Tuskegee, Alabama; Hartford, Connecticut; and San Antonio, Texas. The RDD sample of households used a simple random sample in each of the cities and was partially screened for nonworking or business numbers. The goal was to achieve the completion of 25-minute telephone interviews with 900 adults in the following groups: (1) 300 blacks (100 in Hartford; 100 in Birmingham; 100 in Tuskegee); (2) 400 whites (100 in Hartford; 100 in Birmingham; 100 in Tuskegee; 100 in San Antonio); (3) 100 Hispanics of Puerto Rican descent (Hartford); and (4) 100 Hispanics of Mexican-American descent (San Antonio). The primary sampling unit for the calling area was the Central Office Code, the 3-digit telephone exchange used for local calling areas, that provided a concise area for geographic exposure and increased the likelihood of contacting "households" through RDD.

In 2003, the TLP Questionnaire was administered by ORC Macro, a US-based international opinion research corporation. The survey sample for this study was drawn from the total noninstitutionalized adult populations of persons over the age of 18 residing in telephone-equipped dwelling units in New York, New York; Baltimore, Maryland; and San Juan, Puerto Rico. Each of the 3 sites was sampled independently, with a target number of 900 completed interviews overall and the following specific race/ethnic groups targets within each city: 300 blacks (150 in New York City, 150 in Baltimore); 300 whites (150 in New York City, 150 in Baltimore), and 300 Hispanics (150 Puerto Rican

Hispanics in San Juan, 150 in New York City). The initial sample allocation across strata (within site) was based on expected yields computed using exchange incidence data. The yield of minority interviews by strata was monitored, and the mix of sample across race/ethnic strata in each replicate was adjusted, based on actual data.

Experienced, trained interviewers conducted the survey using full computer-assisted telephone interview (CATI) technology; they were supervised at all times and were randomly electronically monitored for both surveys. The 1999 and 2000 4-city TLP survey study was approved by the institutional review boards of the University of Connecticut and New York University; the 2003 3-city TLP survey study was approved by the institutional review board of New York University. Details regarding the administration of the 1999 and the 2003 TLP surveys, including rationale for selection of the cities, have been previously published.[2,8]

Measures

The TLP Questionnaire addresses a range of issues related to the recruitment of minorities into biomedical studies, specifically, whether minorities are more reluctant to participate in biomedical research studies and, if so, why. A question in the TLP specifically explored the belief that the US government created the human immunodeficiency virus in a laboratory in order to selectively kill a specific group of people ("How likely is it that AIDS is the result of a government plan to intentionally kill a certain group of people by genocide?"). This question occurred two-thirds of the way through the questionnaire (question 42 out of 60) and was prefaced with the statement, "AIDS has killed many people in the United States and throughout the world." Possible categorical responses to each question include "very likely," "somewhat likely," "don't know or not sure," "somewhat unlikely," or "very unlikely." We grouped persons who answered "very likely" or "somewhat likely" together and considered them to be "believers" in the statement and grouped those who responded "very unlikely" or "somewhat unlikely" together as "nonbelievers." In addition, although we examined the distribution by race/ethnicity and gender of those who answered "don't know or not sure," we excluded these persons from further analysis (by coding these respondents as "missing") because we considered them to be neither true believers nor true non-believers. However, because we considered arguments that this group might "act like" either believers or nonbelievers, we tested this assumption by grouping those who stated "don't know or not sure" first with the believers and then with the nonbelievers in order to analyze whether our recategorization appreciably changed the analysis outcome.

To measure 2 constructs, "likelihood of participation" and "fear of participation," 2 scales were created from multiple, related questions using standard psychometric analysis techniques. Both scales, the Likelihood of Participation (LOP) scale and the Guinea Pig Fear Factor (GPFF) scale, were standardized and measured on a continuous scale of 0 to 100, with a score of 0 indicating the lowest willingness to participate and the lowest fear of participation and 100 indicating the greatest willingness to participate and the greatest fear of participation in biomedical research studies. Details regarding the properties of these scales have been previously described.[2]

We calculated age using date of birth, and categorized age as 19 to 34 years old, 35 to 49 years old, or 50 years old or older. We classified education as less than high school graduate, high school alone or with some college, or college graduate and/or postgraduate degree. We categorized income as less than $20,000 a year, equal to or greater than $20,000 but less than $75,000 a year, or greater than or equal to $75,000. To acknowledge and account for cultural differences between the cities (ie, above and beyond simple demographic differences of respondents), the variable of "city" was included as a separate covariate in all multivariate analyses (Hartford, Birmingham, Tuskegee, or San Antonio for the survey conducted in 1999 and 2000 and New York, Baltimore, or San Juan in 2003).

Statistical Analysis

We analyzed each survey separately and then compared the findings of the 2 surveys. Because of the complex design of the 2003 3-city study, it was necessary to develop sampling weights for analysis of these data. Data from this survey were weighted separately in 3 stages, which yielded a

set of analytical weights that were included in the analysis of these survey data to avoid biased estimates. We compared these estimates with analyses without using weights and report whether these estimates changed the results substantially. We tested whether effects were modified by including interaction terms for race/ethnicity with age (older than 35 vs 35 or younger), gender, education, and income in our analysis.

We first examined, descriptively, demographic characteristics of the respondents in each survey by race/ethnic group (using analysis of variance [ANOVA] and chi-square tests) and then explored, for each survey, how the distribution of responses to the question "Is AIDS the result of a government plan to intentionally kill a certain group of people by genocide?" varied by race/ethnic category (white, black, Hispanic). We then evaluated, for each survey, whether a belief in this statement was related to age, gender, education, income, or geographic location within each racial/ethnic category, using Student *t* tests (for age) and chi-square or Fisher exact tests (for gender, education, income, and geographic location). We then conducted logistic regression analyses in which we included only those who responded "likely" (ie, "believers," very or somewhat) and "unlikely" (ie, "nonbelievers," very and somewhat) to examine whether the belief that "AIDS is the result of a government plan to intentionally kill a certain group of people by genocide" was related to race/ethnicity, controlling for differences between the race/ethnic groups in demographics when these variables were found to vary.

To compare the relationship between belief in AIDS origin conspiracy theory and fear of participation (as measured by the GPFF scale) and likelihood of participation (as measured by the LOP scale), we used linear regression to determine whether the dependent variables, GPFF and LOP, were related to the belief that "AIDS is the result of a government plan to intentionally kill a certain group of people by genocide" (dichotomized as yes/no) after first confirming that GPFF and LOP were normally distributed. For each racial/ethnic category, we created a full model with demographic factors (age, gender, education, income, and geographic location) as the independent variables and predictors of the 2 dependent variables, LOP and GPFF, setting the removal value at 0.10.

RESULTS

Study Sample

For both surveys, the original targeted enrollment goals were met or exceeded in each city with the exception of the Puerto Rican group in the 1999 and 2000 survey. Response rates were 70% for Birmingham, 65% for Tuskegee, 49% for Hartford, and 50% for San Antonio in 1999 and 2000, and 52% for San Juan, 51% for Baltimore, and 44% for New York City in 2003. Overall survey completion rates exceeded 90% in 1999 and 2000 and 82% in 2003. The overall final study sample for the 1999 and 2000 survey consisted of 353 blacks, 623 whites, and 157 Hispanics (116 Mexican and 41 Puerto Rican participants were grouped together as Hispanics; total n = 1,133), and the 2003 survey included 356 blacks, 493 whites, and 313 Puerto Rican Hispanics (total n = 1,162).

Demographic characteristics of the overall final study samples for each survey are shown in **Table 1**. Hispanic participants were younger than blacks and whites (ANOVA, $P \leq .001$) in each survey. In the 1999 survey, Hispanics were less likely to be men, compared with whites and blacks ($\chi^2_2 = 7.0$, $P = .03$); in the 2003 survey, there were no differences in gender distribution of the groups (in both the weighted and unweighted analysis). In both surveys, whites reported higher levels of education (1999: $\chi^2_4 = 65.6$, $P \leq .001$; 2003: $\chi^2_4 = 38.0$, $P \leq .001$) and higher incomes (1999: $\chi^2_4 = 80.7$, $P \leq .001$; 2003: $\chi^2_4 = 69.4$, $P \leq .001$) compared with blacks and Hispanics (both in the weighted and unweighted analyses).

Belief in AIDS Origin Conspiracy Theory

The distribution by race/ethnicity of responses in each survey to the question of whether AIDS was "the result of a government plan to intentionally kill a certain group of people by genocide" is shown in **Figure 1**. Patterns of responses by race/ethnic group to the AIDS origin conspiracy question were consistent across surveys. Compared with whites, in the 1999 survey, roughly 3 times as many blacks and Hispanics reported that it was "very or somewhat likely" that AIDS is "the result of a government plan to intentionally kill a certain group of people by genocide" (27.8% of blacks, 23.6% Hispanics vs 8% of whites; $\chi^2_4 = 80.6$, $P \leq .001$). In the 2003 survey, roughly 4 times as

Table 1. Demographic characteristics of study subjects by survey and racial/ethnic group[a]

	4-city TLP survey, 1999 and 2000 (n = 1,133)			3-city TLP survey, 2003 (n = 1,162)		
	Black 31.2 % (n=353)	Hispanic 13.9% (n=157)	White 55.0% (n=623)	Black 31.2% (n=356)	Hispanic 13.9% (n=313)	White 55.0% (n=493)
Mean age, years (±*SD*)	49.1 (± 16.5)	41.6 (± 16.1)	53.9 (± 17.0)	47.1 (± 15.5)	44.3 (± 15.8)[a]	48.2 (± 17.1)
Gender, % men	52.1%	39.5%	48.3%	32.6%	31.6%	36.7%
Education						
< High school graduate	21.5%	26.1%	11.7%	18.0%	21.7%	11.8%
High school graduate and/or some college	60.4%	64.3%	51.1%	53.7%	40.9%	42.0%
College graduate and/or postgraduate degree	17.9%	15.3%	36.8%	27.8%	36.7%	45.6%
Family income (yearly)						
<$19,999	40.2%	36.9%	18.5%	31.5%	38.7%	18.1%
$20,000 to 74,999	49.0%	46.5%	50.7%	54.2%	45.4%	48.1%
≥$75,000	4.8%	5.1%	17.7%	8.1%	7.3%	20.5%
Geographic location						
Hartford, CT	36.0%	23.6%	35.6%	NA	NA	NA
Birmingham, AL	29.5%	0.6%	16.9%	NA	NA	NA
Tuskegee, AL	30.0%	0.6%	29.5%	NA	NA	NA
San Antonio, TX	4.5%	75.2%	18.0%	NA	NA	NA
New York, NY	NA	NA	NA	55.9%	2.2%	33.3%
Baltimore, MD	NA	NA	NA	47.2%	47.9%	63.7%
San Juan, PR	NA	NA	NA	2.2%	49.8%	3.0%

Note: TLP = Tuskegee Legacy Project; NA = not applicable.

[a] Values may not add to 100% because of small amounts (<5%) of missing data.

many blacks and almost 3 times as many Hispanics compared with whites (34.1% of blacks, 21.9% of Hispanics vs 8.4% whites; $\chi^2_4 = 80.6$, $P \leq .001$) reported that it was "very or somewhat likely" that AIDS is "the result of a government plan to intentionally kill a certain group of people by genocide." In the 2003 survey weighted analysis, 4.1% of whites, 5.6% of Puerto Rican Hispanics, and 9.5% of blacks reported belief in AIDS origin conspiracy theory (Rao-Scott $\chi^2_4 = 61.3$, $P \leq .001$).

The proportion of persons who responded "don't know or not sure" to this question was greater in the 2003 TLP survey (range, 14.1% to 32.2%) compared with the 1999 TLP survey (range, 6.1% to 12.2%). However, we chose the more conservative analysis of excluding this group from further analyses in both surveys, because including this group in the group of "believers" or in the group of "nonbelievers" did not reveal consistent demographic or other patterns regarding this group of persons, and we hypothesized that those who responded "don't know/not sure" were not committed to either side (ie, believing or not believing in AIDS origin conspiracy theory).

When controlling for differences in age, gender, education, income, and geographic location, we found that, compared with whites, blacks and Hispanics were still much more likely to report belief in AIDS origin conspiracy theory (1999: blacks, adjusted odds ratio (OR) 3.1; 95% CI, 2.2–4.2; Hispanics adjusted OR 3.2; 95% CI, 1.9–5.3; 2003: blacks adjusted OR 7.4; 95% CI, 4.8–11.4; Hispanics

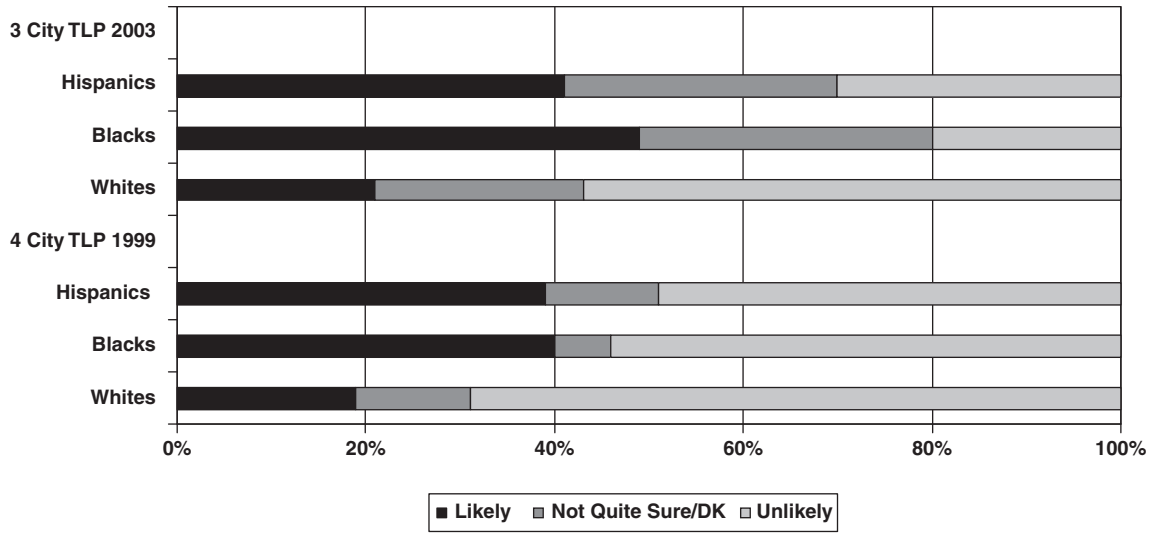


Figure 1. AIDS origin beliefs, by race/ethnicity, across 2 Tuskegee Legacy Project (TLP) Questionnaire surveys (1999 and 2003). DK = don't know.

adjusted OR 3.3; 95% CI, 1.9–5.7). We found no evidence of effect modification as no interaction terms were significant in our analysis.

Demographic Characteristics and Belief in Conspiracy Theory

The relationships between demographic factors (age, gender, education, income, and geographic location) and belief in AIDS origin conspiracy theory (based on response to the question on intentional killing and genocide) within each racial/ethnic group are shown in **Table 2**. Gender and age were not related to being a believer in any racial/ethnic subgroup in either survey. In general, lower education and lower income levels were related to a belief in AIDS origin conspiracy within each racial/ethnic group, with the exception of income in Hispanics and education in blacks in the 1999 survey. Geographic location was not related to being a believer in any racial/ethnic with the exception of Puerto Rican Hispanics in the 2003 survey, where those living in New York were more likely to express belief in AIDS origin conspiracy theory. These results were confirmed by the weighted analysis in the 2003 survey.

Relationship of Belief in AIDS Origin Conspiracy Theory, Likelihood of Participation, and Fear of Participation

We found no significant associations in whites, blacks, or Hispanics between belief in AIDS origin conspiracy theory and GPFF or LOP in either the 1999 or the 2003 TLP surveys (including the weighted analysis for the 2003 survey). We did identify a consistent relationship (ie, a significant association in both 1999 and 2003) between belief in AIDS origin conspiracy theory and an increased fear of participation in biomedical research (as measured by GPFF score) among whites when we adjusted for income and education (**Table 3**); this was confirmed by the weighted 2003 analysis. No consistent relationships were found within the black or Hispanic groups in adjusted analyses.

DISCUSSION

We found across 2 large surveys spaced 3 to 4 years apart that blacks and Hispanics were more likely than whites to report a belief in AIDS origin conspiracy theory. However, we found that for the racial/ethnic minority groups, holding such a belief was not associated with an expressed decreased likelihood of participation in, or an expressed increased fear of participation in,

Table 2. Demographic predictors of belief that "AIDS is the result of a government plan to intentionally kill a certain group of people by genocide" held by whites, blacks, and Hispanics across 2 studies

	4-city TLP, 1999 and 2000			3-city TLP, 2003		
	Blacks	Hispanics	Whites	Blacks	Hispanics	Whites
Age[a]	NS	NS	NS	NS	NS	NS
Gender[b]	NS	NS	NS	NS	NS	NS
Education[b]	NS	Lower education more likely to believe (P = .06)	Lower education more likely to believe (P = .005)	Lower education more likely to believe (P = .09)	Lower education more likely to believe (P = .04)	Lower education more likely to believe (P = .001)
Income[b]	Lower income more likely to believe, (P = .001)	NS	Lower income more likely to believe (P = .02)	Lower income more likely to believe (P = .06)	Lower income more likely to believe (P = .03)	Lower income more likely to believe (P = .001)
City[b]	NS	NS	NS	NS	Hispanics in NY more likely to believe (P = .002)	NS

Note: TLP = Tuskegee Legacy Project; NS = nonsignificant.
[a]ANOVA.
[b]Chi-square test.

biomedical research. The results of this investigation concur with many other reports that have demonstrated that AIDS conspiracy beliefs are more commonly held by blacks than by whites in the United States,[13–15,19–28,30] and they add to the growing literature that these beliefs are likely more common in other minority racial/ethnic groups, including Hispanics. A 1990 poll by *The New York Times* reported that 29% of blacks but only 13% of whites in New York City agreed with the statement "the virus which causes AIDS was deliberately created in a laboratory in order to infect Black people."[30] Herek reported that twice as many blacks as whites felt the government "is not telling the whole story about AIDS"[28] and reported that 5 times as many blacks as whites agreed with the statement "the government is using AIDS to kill off minority groups."[27] Thomas showed that anywhere from 15% to 35% of respondents agreed with the statement "I believe AIDS is a form of genocide against the Black race,"[29] and Goertzel reported that 31% of blacks in Philadelphia area believed that "the government deliberately spread the AIDS virus in the black community."[26] Klonoff reported that 26.5% of blacks agreed with the statement that "HIV/AIDS is a man-made virus that the federal government made to kill and wipe out Black people,"[24] and a group of studies by Bogart reported that 70% of blacks agreed with the statement "a lot of information about AIDS is being held back from the public," 50% with the statement that "HIV is a manmade virus," and 26% with the statement "AIDS is a form of genocide against African Americans."[14,15] More recent studies confirm that AIDS conspiracy beliefs continue to be persistent in the black community.[13,19–21,31] To date, only 4 studies measuring the prevalence of AIDS conspiracy theories have included data on Hispanics, the fastest growing minority group in the United States and one that is, like blacks, disproportionally affected by AIDS/HIV.[13,21–22,26] An attempt to identify barriers to participation in clinical trials among Hispanics is important, given the disparities in HIV rates by ethnicity and their underrepresentation in biomedical research, including HIV vaccine trials.[33] Our finding that

Table 3. Relationship of belief in conspiracy theories and likelihood of participation and fear of participation in whites, blacks, and Hispanics across 2 studies[a]

	4-city TLP, 1999 and 2000			3-city TLP, 2003		
Outcome	Whites	Blacks	Hispanics	Whites	Blacks	Hispanics
Likelihood of participation						
Age	NS	NS	NS	NS	−0.14 (.02)	NS
Gender	NS	NS	NS	NS	NS	NS
Education	NS	−0.17 (.006)	NS	NS	NS	−0.21 (.002)
Income	NS	−0.18 (.004)	NS	NS	NS	NS
City of residence	NS	NS	NS	NS	NS	0.19 (.007)
Belief in conspiracy theories	**0.08 (.06)**	**NS**	**0.19 (.03)**	**NS**	**NS**	**NS**
Fear of participation						
Age	NS	−0.16 (.005)	NS		−0.21 (.001)	−0.14 (.005)
Gender	NS	0.11 (.06)	NS	NS	NS	NS
Education	NS	NS	NS	−0.12 (.04)	NS	NS
Income	NS	NS	NS	−0.10 (.08)	NS	NS
City of residence	NS	NS	NS	NS	NS	NS
Belief in conspiracy theories	**0.09 (.04)**	**NS**	**NS**	**0.13 (.02)**	**NS**	**NS**

Note: TLP = Tuskegee Legacy Project; NS = nonsignificant.

[a]Values are given as standardized β-coefficients (*P* value). Standardized regression coefficients were created using linear regression. Criteria for entry: probability of t=0.

AIDS origin conspiracy theory belief is more common among Hispanics compared with whites concurs with 3 of these studies.[13,22,26]

We used the same methodology and the same wording in 1999 and 2000 and in 2003, and we found that results across the 2 surveys were similar, although the proportion of blacks reporting agreement with our AIDS conspiracy statement was slightly higher in the more recent survey (34.1% vs 27.8% in the earlier survey). We also found that a higher proportion of persons of any racial/ethnic category reported that they were "not quite sure" in 2003 compared with 1999 and 2000 and that the adjusted odds ratios for belief in AIDS origin conspiracy theories were twice as high for blacks as for whites in 2003 versus 1999. Differences in these rates may be related to the fact that the cities sampled were different or may be due to a true secular trend, which could be related to events occurring after the first but before the second survey (eg, the September 11th attacks of 2001) or to the government administration (ie, the election of George W. Bush following Bill Clinton's term as president). The latter may have caused some in the black community, who are statistically more likely to be members of the Democratic Party, to feel a greater disenfranchisement from the US government.

Although other surveys have found that conspiracy theory beliefs are more common in men, we failed to identify statistically significant gender differences within any racial/ethnic group. However, this study is likely to have limited power to identify differences between genders that were not

large. Regarding issues of education and income, our results concur with other reports that show that lower income and education are related to belief in AIDS origin conspiracy theory.[14,15] Because we adjusted for differences in education and income among the racial/ethnic groups, we are confident that the increase in belief in AIDS origin conspiracy theory is likely related to racial/ethnic categorization and not to differences in income, education, or other factors between groups.

To date, despite the fact that "minorities, overrepresented in the HIV epidemic are underrepresented in HIV clinical trials,"[34] only one other study has investigated whether belief in AIDS conspiracy theories is related to participation in biomedical research.[31] This investigation, which explored racial/ethnic differences in knowledge of and willingness to participate in HIV vaccine trials, was a cross-sectional survey of a convenience sample of 220 students attending a community college in Atlanta, Georgia. Students were asked a series of questions regarding beliefs of AIDS, including "the AIDS epidemic was caused by a government conspiracy." Willingness to participate in an HIV vaccine trial was measured by asking respondents the following question: "After a researcher told me the details of an HIV vaccine clinical trial, I would enroll in one." In that study, 25% of blacks versus 12% of whites stated that AIDS was caused by a government conspiracy; willingness to participate in an HIV vaccine clinical trial did not differ by race and was not associated with belief that AIDS was caused by a government conspiracy. Our survey results are remarkably similar despite the difference in sampling methods used. Even though we also identified a widespread belief that AIDS is "the result of a government plan to intentionally kill a certain group of people by genocide," the expressed likelihood of participation and fear of participation were not consistently associated with belief in AIDS origin conspiracy theory in blacks; in whites, only the expressed likelihood of participation was related to this belief. That earlier study, however, did not include Hispanics and included respondents drawn from a narrow segment of the population (community college students in a southern American city).

We found that among whites, the belief that AIDS is "the result of a government plan to intentionally kill a certain group of people by genocide" was directly related to fear of participation in biomedical research. Again, it is possible that because in each survey more participants were white, we had greater power to sense an effect of belief in AIDS origin conspiracy theory. The overall goal of the TLP, however, was to determine whether there is a greater reluctance to participate in biomedical studies among minorities as compared with whites and, if so, to explore the factors that might account for that observed difference. Attainment of this goal is critical to ensure that the findings from biomedical studies provide health data on the diverse populations of the United States, to assist biomedical researchers in achieving compliance with current 1994 National Institutes of Health (NIH) Guidelines for the Inclusion of Women and Minorities in Clinical Studies, and to provide empirical suggestions for intervention studies on enlisting minorities into biomedical studies including clinical trials.

Identification of barriers to recruitment of racial and ethnic minorities is an essential step in creating greater ethnic/racial balance in biomedical research studies. It appears that the belief that AIDS origin conspiracy theory is widely held among racial and ethnic minority groups, but this belief does not, according to our results, appear to be related to self-reported willingness to participate in biomedical research in these population subgroups. Indeed, we found such a relationship in whites but not in either of the 2 minority groups. Although this study had greater power to detect differences among whites because of a greater number of white participants in both the 1999 and the 2003 TLP surveys, differences in the GPFF and LOP scales among blacks and Hispanics were small and unlikely to be of clinical significance. Indeed, major advantages of this survey were its focus on comparisons among racial/ethnic groups and, along with its carefully worded and extensively tested questionnaire, its sampling methodology and large sample size.

As in any telephone survey, we are unable to make conclusions about persons who chose not to respond. It is possible that those who did not respond to our questionnaire were no different from those who responded. However, it is possible that response bias may have an effect on our absolute numbers, for example, the proportion in each group might have been lower or higher than reported. However, we believe it is unlikely that persons from various racial/ethnic groups would

have refused differentially to take part in the study, and therefore we believe the effect of nonresponse bias in the interpretation of the degree of differences across the groups is of only minor concern.

CONCLUSIONS

We found that blacks and Hispanics were more than 3 times as likely as whites to be believers in this AIDS origin conspiracy theory but that holding such a belief was not associated with a decreased likelihood of participation in, or increased fear of participation in, biomedical research.

ACKNOWLEDGMENTS

This study was supported within the New York Universtiy Oral Cancer RAAHP (Research on Adolescent and Adult Health Promotion) Center, a National Institute of Dental and Craniofacial Research /National Institutes of Health Oral Health Disparities grant (U54 DE014257).

REFERENCES

1. Swanson MS, Ward AJ. Recruiting minorities into clinical trials: toward a participant-friendly system. *J Natl Cancer Inst.* 1995;87(23):1747–1759.
2. Katz RV, et al. The Tuskegee Legacy Project: willingness of minorities to participate in biomedical research. *J Health Care Poor Underserved.* 2006;17(4):698–715.
3. Sullivan PS, McNaghten AD, Begley E, Hutchinson A, Cargill VA. Enrollment of racial/ethnic minorities and women with HIV in clinical research studies of HIV medicines. *J Natl Med Assoc.* 2007;99(3):242–250.
4. Corbie-Smith GM, Thomas SB, Williams MV, et al. Attitudes and beliefs of African Americans toward participation in medical research. *J Gen Intern Med.* 1999;14: 537–546.
5. Shavers-Hornaday VL, Lynch CF. Why are African Americans under-represented in medical research studies? Impediments to participation. *Ethnic Health.* 1997;2: 31–45.
6. Fouad MN, Corbie-Smith G, Curb D, et al. Special populations recruitment for the Women's Health Initiative: successes and limitations. *Control Clin Trial.* 2004;25: 335–352.
7. Washington, H. *Medical Apartheid: The Dark History of Medical Experimentation on Black Americans from Colonial Times to the Present.* New York: Doubleday; 2007.
8. Katz RV, Green BL, Kressin NR, Claudio C, Wang MQ, Russell SL. Willingness of minorities to participate in biomedical studies: confirmatory findings from a follow-up study using the Tuskegee Legacy Project Questionnaire. *J Natl Med Assoc.* 2007;99(9):1050–1062.
9. Bates BR, Harris TM. The Tuskegee Study of untreated syphilis and public perceptions of biomedical research: a focus group study. *J Natl Med Assoc.* 2004;96(8): 1051–1064.
10. Freimuth VS, Quinn SC, Thomas SB, et al. African Americans' views on research and the Tuskegee Syphilis Study. *Soc Sci Med.* 2001;52:245–253.
11. Shavers VL, Lynch CF, Burmeister LF. Knowledge of the Tuskegee Study and its impact on willingness to participate in medical research studies. *J Natl Med Assoc.* 2000;92(12):563–572.
12. Green BL, Maisiak R, Wang MQ. Participation in health education, health promotion, and health research by African Americans: effects of the Tuskegee Syphilis Experiment. *J Health Educ.* 1997;28(4):196–201.
13. Ross MW, Essien EJ, Torres I. Conspiracy beliefs about the origin of AIDS in four racial/ethnic groups,. *J Acquir Immune Defic Syndr.* 2006;41:342–344.
14. Bogart LM, Thorburn S. Are AIDS conspiracy beliefs a barrier to HIV prevention among African Americans? *J Acquir Immune Defic Syndr.* 2005;38:213–218.
15. Bogart LM, Bird ST. Exploring the relationship of conspiracy beliefs about AIDS to general behaviors and attitudes among African-American adults. *J Natl Med Assoc.* 2003;95(11):1057–1065.
16. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA.* 2004;291(22):2720–2726.
17. Gifford AL, Cunningham WE, Heslin KC. Participation in research and access to experimental treatments by HIV-infected patients. *New Engl J Med.* 2002;346:1373–1382.
18. Sengupta S, Strauss RP, DeVellis R, et al. Factors affecting African-American participation in AIDS research. *J Acquir Immune Defic Syndr.* 2000;24(3):275–284.
19. Zekeri AA, Habtermariam T, Tameru B, Ngawa D, Robnett V. Conspiracy beliefs about HIV/AIDS among HIV-positive patients in rural Alabama. *Psychol Rep.* 2009;104(2):388–394.
20. Bohnert AS, Latkin CA. HIV testing and conspiracy beliefs regarding the origins of HIV among African Americans. *AIDS Patient Care STDS.* 2009;23(9):759–763.
21. Clark A, Mayben JK, Hartman C, Kallen MA, Giordano TP. Conspiracy beliefs about HIV infection are common but are not associated with delayed diagnosis or adherence to care. *AIDS Patient Care STDS.* 2008;22(9):753–759.
22. Hutchinson AB, Begley EB, Sullivan P, Clark HA, Boyett BC, Kellerman SE. Conspiracy beliefs and trust in information about HIV/AIDS among minority men who have sex with men. *J Acquir Immune Defic Syndr.* 2007;45(5): 603–605.
23. Bogart LM, Thorburn S. Relationship of African American's sociodemographic characteristics to belief in conspiracies about HIV/AIDS and birth control. *J Natl Med Assoc.* 2006;98(7):1144–1150.
24. Klonoff EA, Landrine H. Do blacks believe that AIDS is a government conspiracy against them? *Prev Med.* 1999;28(5):451–457.

25. Quinn SC. Belief in AIDS as a form of genocide: implications for HIV prevention programs for African Americans. *J Health Educ.* 1997;28(6):S6-S11.
26. Goertzel T. Belief in conspiracy theories. *Political Psychol.* 1994;15(4):731–742.
27. Herek GM, Capitanio JP. Conspiracies, contagion, and compassion: trust and public reactions to AIDS. *AIDS Educ Prev.* 1994;6(4):365–375.
28. Herek GM, Glunt EK. AIDS-related attitudes in the US: a preliminary conceptualization. *J Sex Res.* 1991;28: 99–123.
29. Thomas SB, Quinn SC. The burdens of race and history on black Americans' attitudes toward needle exchange policy to prevent HIV disease. *J Public Health Policy.* 1993;14(3): 320–347.
30. DeParle J. Talk grows of government being out to get blacks. *New York Times*. October 29, 1990.
31. Priddy FH, Cheng AC, Salazar LF, Frew PM. Racial and ethnic differences in knowledge and willingness to participate in HIV vaccine trials in an urban population in the Southeastern US. *Int J STD AIDS.* 2006;17(2):99–102.
32. Hall HI, Song R, Rhodes P, et al. Estimation of HIV incidence in the United States. *JAMA.* 2008;300:520–529.
33. Sobieszcyk ME, Xu G, Goodman K, Lucy D, Koblin BA. Engaging members of African American and Latino communities in preventive HIV vaccine trials. *J Acquir Immune Defic Syndr.* 2009;51(2):194–201.
34. Cargill VA, Stone VE. AIDS: a minority health issue. *Med Clin North Am.* 2005;89(4):895–912.

EDITOR'S CHOICE

Editorial

Worldwide evidence-based medicine activities

Richard Saitz

10.1136/ebm1172

Editor, *Evidence-based Medicine*

Section of General Internal Medicine, Boston Medical Center, Boston University Schools of Medicine and Public Health, Boston, Massachusetts, USA

Correspondence to
Richard Saitz
Section of General Internal Medicine, Boston Medical Center, Boston University Schools of Medicine and Public Health, 801 Massachusetts Avenue, 2nd floor (Crosstown 2), Boston, MA 02118–2335, USA;
rsaitz@bmjgroup.com

Published Online First
12 January 2011

Evidence-Based Medicine (EBM) serves at least two major communities – those teaching and studying evidence-based medicine (EBM) itself (and producing evidence), and those consuming evidence to improve patient care. Two recent developments of interest to both audiences are worth noting.

First, on 1 and 2 November, 2010, the *British Medical Journal (BMJ)* Evidence Centre and the Centre for Evidence-Based Medicine, Oxford University held an international conference ('evidence2010 Transforming Healthcare') in London. Although this was not a conference of a particular professional society nor was it an annual event (though it may be in the future), the conference was well-attended by a wide range of evidence producers and consumers from around the world. As a whole the participants had refreshing views about evidence-based medicine (EBM). They were neither 'for it' nor 'against it.' They understood that evidence-based medicine (EBM) was complex, has its place, has promised much and has delivered on much of that promise, but also faces many challenges, particularly in the area of how best to apply evidence in practice.

Many of the discussions were quite memorable, and I will mention only a few here (also see http://www.evidence2010.com). The editors of the *Lancet* and *BMJ* spoke, and debated, the future of medical publishing. One prediction was the greater accessibility of medical evidence, particularly online. That future, along with greater registration of clinical trials, will be welcome, because as was pointed out by another speaker, systematic reviews can be misleading when studies are done but never published. Victor Montori discussed the complicated work of being a patient with several chronic diseases and the challenges of applying evidence for specific diseases and interventions when caring for such patients. Although there was much discussion about how evidence is often not translated into practice, Bill Summerskill, when talking about a study that eventually leads to practice change, said "you may never actually be sure when that seed is planted and when it blossoms." That observation is one for implementation researchers to confirm and learn from. I cannot convey the genuine excitement that was evident among conference participants that they were taking part in an important event in the history of evidence-based medicine (EBM).

The other activity relevant for evidence-based medicine (EBM) as a field is the establishment of an international organisation called the International Society for Evidence-based Health Care. The mission is to develop and encourage research in evidence-based healthcare and to promote and provide professional and public education in the field. The society published its first newsletter (available at http://ebm.mcmaster.ca/), and it plans to hold meetings and disseminate evidence-based healthcare education via workshops and educational materials, and to promote research. It will no doubt be of interest to see how the society coalesces and finds its place in the world of professional societies.

To me, these developments indicate that despite, or perhaps because of the many challenges faced by evidence-based medicine (EBM) practitioners, teachers and researchers, evidence-based medicine (EBM) is alive, well and even thriving. I expect *EBM* (the journal) will continue to serve this community well, and I welcome input from you on how it can best do so.

Addiction
EDITORIAL
doi:10.1111/j.1360-0443.2011.03523.x

Russia and human immunodeficiency virus—beyond crime and punishment

Russia faces formidable challenges in the second decade of its major human immunodeficiency virus (HIV) epidemic. Although Russian public health and medical responses are growing, they remain inadequate for the current crisis. HIV prevention lessons can be learnt by Russia from mistakes and innovations of the early epidemic in the United States, western Europe and Australia.

The human immunodeficiency virus (HIV) epidemic in Russia has been driven by injection drug use. The dire consequences of this epidemic in Russia reflect their public health and medical care approach to both HIV infection and drug users. Many of the mistakes that were made in the United States in the first two decades of the acquired immune deficiency syndrome (AIDS) epidemic with regard to HIV infection and drug users are being repeated in Russia at a comparable time-period in the disease's evolution in that country. Criticism has been leveled at Russia for its lack of adherence to the World Health Organization's recommendations for preventing HIV in drug users [1–3]. Two decades of additional experience with HIV among drug users in other countries may provide insight to advance successes and mitigate failures in Russia's second decade of the HIV epidemic.

Russia has approximately 1 million HIV-infected people among its 140 million population [4]. Eighty-three per cent have a history of injection drug use and most of the remainder are sexual partners of injection drug users (IDUs) [4]. The estimated HIV incidence rate among IDUs in St Petersburg is 14.1 per 100 person-years [5]. Simply stated, the HIV epidemic is not under control. Given that 1.8 million IDUs currently live in Russia [2], and that condom use by Russian drug users is not the norm [6], there is great risk for further expansion of the epidemic.

In the United States and Europe, the clinical syndrome that became known as AIDS resulting from HIV infection was spreading actively among IDUs during the late 1970s and early 1980s. It was not a problem in Russia at that time. HIV only began to spread among IDUs analogously in Russia in epidemic proportions two decades later, in the late 1990s and at beginning of the new millennium [7].

During the first two decades of the AIDS epidemic, mistakes were made in the United States and some other countries with regard to the public health and medical response; hence, HIV seroprevalence in IDUs was high, ranging from 22% to more than 50% [8]. Exceptions to this trend exist, most notably in Australia with an HIV seroprevalence rate among IDUs of 1% [9]. This is due largely to that country's aggressive adoption of harm reduction measures, such as syringe exchange and opioid agonist therapy (OAT) programs, as early as the mid-1980s. In the United States, distribution of clean needles and syringes to IDUs was shunned by government officials with disastrous consequences. Addiction treatment and OAT, essentially methadone, existed but were not actively expanded and improved in order to confront the epidemic directly (e.g. routine implementation of HIV testing within methadone programs was not instituted initially). More than a decade after the onset of the US/western European AIDS epidemic, while advancements were being achieved in elucidating viral etiology and developing effective antiretroviral medications, researchers reported that an infected person's entry into HIV care was commonly not occurring until the disease had become advanced [10,11]. Active efforts to engage HIV-infected people and, in particular, HIV-infected drug and alcohol users, were inadequate and not a high priority on the HIV activists or researchers' agenda.

Today in Russia, two of the most glaring policy mistakes from a public health perspective are the absence of OAT and the suboptimal distribution of clean needles and syringes [1]. Without a doubt, these measures have proved themselves immensely valuable in limiting the spread of infection even in the face of ongoing drug use. Needle availability from pharmacies in much of Russia is a step in the right direction, but insufficient to meet the needs of HIV prevention efforts.

One very positive dimension of HIV prevention among drug users in Russia is the comprehensive 'opt-out' strategy for HIV testing that has long existed in many spheres, particularly within the narcology treatment system [12]. Such a phenomenon differs sharply with the approach that existed in a comparable period in the United States, during which unawareness of HIV status was the major factor underlying the years of delay between HIV infection and entry into medical care [13].

Despite these testing policies, HIV-infected IDUs in Russia are not receiving care at acceptable rates. Although drug use accounts for more than 80% of Russian HIV infections, IDUs comprise fewer than half of those receiving HIV care [14]. In fact, Russia is among the countries with the lowest antiretroviral therapy (ART) coverage for those with advanced HIV [4].

One source of delay into medical care for those with HIV infection is the time between awareness of HIV infection and engaging with HIV medical care [15]. This is a poorly documented but suspected problem faced by many HIV-infected drug users in Russia in 2011. While IDUs face barriers to HIV care such as stigma, discrimination, addiction-related priorities, providers' pessimistic perceptions regarding adherence and unstable living situations [16,17], the problem is also structural. Although the HIV care system is evolving with introduction of narcologists into some AIDS treatment centers, coordination of narcology and HIV care remains unaddressed. Medical care in Russia is provided within silos of excellence with specific expertise. In a narcology hospital, addiction is addressed. In an infectious disease hospital, infections such as HIV and hepatitis B and C are treated. However, coordination of care between different physical sites of care is exceedingly limited, and yet patients who suffer from both maladies, addiction and HIV, are common in Russia. Those affected do not benefit from the expertise of providers of both disciplines giving coordinated medical care. Such fragmented systems of care have been shown to worsen access to HIV treatment [18]. It is time to actively facilitate thousands of infected Russian substance users into HIV care; with appropriate systems established, many could transition directly from narcology treatment.

Russia's long-standing excellent record of HIV testing is a bright light within the country's strategic plan to address HIV infection, but it is not sufficient. Innovative strategies need to be employed to engage individuals into the HIV care system. What will it take to engage HIV-infected drug users into care? Here, lessons from those countries with the unfortunate additional two decades of history addressing HIV among drug users could be informative.

Enhanced engagement of drug users into HIV care will be advanced by addressing five major challenges: reducing stigma towards drug users and the HIV-infected; developing protocols to facilitate transitions and coordination between HIV and narcology treatment systems; developing multiple points of entry for individuals into HIV and addiction care; expanding needle and syringe exchanges; and adopting OAT to treat opioid dependence with pharmacological agents. Engagement will enable efforts to achieve effective delivery of antiretroviral treatment to drug users. Engagement will also provide opportunities to mitigate sexual risk behaviors and its consequences among drug users, a high priority as the epidemic spreads from IDUs to the general Russian population [4]. Such efforts could focus on unsafe sex behaviors [19] and treatment of sexually transmitted diseases.

In the first two decades of the HIV epidemic in the United States, many missed opportunities to confront this disease effectively among drug users occurred. Bringing that experience to bear in a very different and yet not so different context holds opportunity to help transform the raging Russian HIV epidemic into one that employs the best strategies to engage reluctant patients in medical care and reduce the spread and consequences of this 21st-century Russian scourge.

Declaration of interests

None.

JEFFREY H. SAMET

Boston University School of Medicine/Boston Medical Center, Department of Medicine, Section of General Internal Medicine, Clinical Addiction Research and Education (CARE) Unit, Boston, MA, USA and Boston University School of Public Health, Department of Community Health Sciences, Boston, MA, USA.

E-mail: jsamet@bu.edu

References

1. Holt E. Russian injected drug use soars in face of political inertia. *Lancet* 2010; **376**: 13–4.
2. Mathers B. M., Degenhardt L., Adam P., Toskin I., Nashkhoev M., Lyerla R. *et al*. Estimating the level of HIV prevention coverage, knowledge and protective behavior among injecting drug users: what does the 2008 UNGASS reporting round tell us? *J Acquir Immune Defic Syndr* 2009; **52**: S132–S142.
3. World Health Organization (WHO), United Nations Office on Drugs and Crime (UNODC), Joint United Nations Programme on HIV/AIDS (UNAIDS). *Technical guide for countries to set targets for universal access to HIV prevention, treatment and care for injecting drug users*. 2009. Available at: http://www.unodc.org/documents/hiv-aids/idu_target_setting_guide.pdf (accessed 31 March 2011; archived by Webcite at http://www.webcitation.org/5zkuSrlOw).
4. Joint United Nations Programme on HIV/AIDS (UNAIDS). *Global Report: UNAIDS Report on the Global AIDS Epidemic, 2010*. Joint United Nations Programme on HIV/AIDS. 2010. Available at: http://www.unaids.org/documents/20101123_GlobalReport_em.pdf (accessed 17 May 2011; archived by Webcite at http://www.webcitation.org/5zkuXN3dx).
5. Niccolai L. M., Verevochkin S. V., Toussova O. V., White E., Barbour R., Kozlov A. P. *et al*. Estimates of HIV incidence among drug users in St. Petersburg, Russia: continued growth of a rapidly expanding epidemic. *Eur J Public Health* 2010; Epub ahead of print 26 August.
6. Niccolai L. M., Shcherbakova I. S., Toussova O. V., Kozlov A. P., Heimer R. The potential for bridging of HIV transmission in the Russian Federation: sex risk behaviors and HIV prevalence among drug users (DUs) and their non-DU sex partners. *J Urban Health* 2009; **86**: s131–s143.
7. Krupitsky E., Zvartau E., Karandashova G., Horton N. J., Schoolwerth K. R., Bryant K. *et al*. The onset of HIV

infection in the Leningrad region of Russia: a focus on drug and alcohol dependence. *HIV Med* 2004; **5**: 30–3.

8. Karon J. M., Fleming P. L., Steketee R. W., De Cock K. M. HIV in the United States at the turn of the century: an epidemic in transition. *Am J Public Health* 2001; **91**: 1060–8. PMCID:PMC1446722.
9. Wodak A., Maher L. The effectiveness of harm reduction in preventing HIV among injecting drug users. *NSW Public Health Bull* 2010; **21**: 69–73.
10. Porter K., Wall P. G., Evans B. G. Factors associated with lack of awareness of HIV infection before diagnosis of AIDS. *BMJ* 1993; **307**: 20–3. PMCID:PMC1678470.
11. Samet J. H., Retondo M. J., Freedberg K. A., Stein M. D., Heeren T., Libman H. Factors associated with initiation of primary medical care for HIV-infected persons. *Am J Med* 1994; **97**: 347–53.
12. Mounier-Jack S., Nielsen S., Coker R. J. HIV testing strategies across European countries. *HIV Med* 2008; **9**: 13–19.
13. Samet J. H., Freedberg K. A., Savetsky J. B., Sullivan L. M., Stein M. D. Understanding delay to medical care for HIV infection: the long-term non-presenter. *AIDS* 2001; **15**: 77–85.
14. Bobrova N., Alcorn R., Rhodes T., Rughnikov I., Neifeld E., Power R. Injection drug users' perceptions of drug treatment services and attitudes toward substitution therapy: a qualitative study in three Russian cities. *J Subst Abuse Treat* 2007; **33**: 373–8.
15. Samet J. H., Freedberg K. A., Stein M. D., Lewis R., Savetsky J., Sullivan L. *et al*. Trillion virion delay: time from testing positive for HIV to presentation for primary care. *Arch Intern Med* 1998; **158**: 734–40.
16. Amirkhanian Y. A., Kelly J. A., Kuznetsova A. V., Difranceisco W. J., Musatov V. B., Pirogov D. G. People with HIV in HAART-era Russia: transmission risk behavior prevalence, antiretroviral medication-taking, and psychosocial distress. *AIDS Behav* 2010; **15**: 767–77.
17. Wood E., Kerr T., Tyndall M. W., Montaner J. S. A review of barriers and facilitators of HIV treatment among injection drug users. *AIDS* 2008; **22**: 1247–56.
18. Tkatchenko-Schmidt E., Atun R., Wall M., Tobi P., Schmidt J., Renton A. Why do health systems matter? Exploring links between health systems and HIV response: a case study from Russia. *Health Policy Plan* 2010; **25**: 283–91.
19. Samet J. H., Krupitsky E. M., Cheng D. M., Raj A., Egorova V. Y., Levenson S. *et al*. Mitigating risky sexual behaviors among Russian narcology hospital patients: the PREVENT (Partnership to Reduce the Epidemic via Engagement in Narcology Treatment) randomized controlled trial. *Addiction* 2008; **103**: 1474–83. PMCID:PMC2588416.

The Impact of a Novel Computer-Based Decision Aid on Shared Decision Making for Colorectal Cancer Screening: A Randomized Trial

Paul C. Schroy III, MD, MPH, Karen Emmons, PhD, Ellen Peters, PhD, Julie T. Glick, MPH, Patricia A. Robinson, MEd, Maria A. Lydotes, BS, Shamini Mylvanaman, MPH, Stephen Evans, MPH, Christine Chaisson, MPH, Michael Pignone, MD, MPH, Marianne Prout, MD, MPH, Peter Davidson, MD, Timothy C. Heeren, PhD

Background. *Eliciting patients' preferences within a framework of shared decision making (SDM) has been advocated as a strategy for increasing colorectal cancer (CRC) screening adherence. Our objective was to assess the effectiveness of a novel decision aid on SDM in the primary care setting.* ***Methods.*** *An interactive, computer-based decision aid for CRC screening was developed and evaluated within the context of a randomized controlled trial. A total of 665 average-risk patients (mean age, 57 years; 60% female; 63% black, 6% Hispanic) were allocated to 1 of 2 intervention arms (decision aid alone, decision aid plus personalized risk assessment) or a control arm. The interventions were delivered just prior to a scheduled primary care visit. Outcome measures (patient preferences, knowledge, satisfaction with the decision-making process [SDMP], concordance between patient preference and test ordered, and intentions) were evaluated using prestudy/poststudy visit questionnaires and electronic scheduling.* ***Results.*** *Overall, 95% of patients in the intervention arms identified a preferred screening option based on values placed on individual test features. Mean cumulative knowledge, SDMP, and intention scores were significantly higher for both intervention groups compared with the control group. Concordance between patient preference and test ordered was 59%. Patients who preferred colonoscopy were more likely to have a test ordered than those who preferred an alternative option (83% v. 70%;* $P < 0.01$*). Intention scores were significantly higher when the test ordered reflected patient preferences.* ***Conclusions.*** *Our interactive computer-based decision aid facilitates SDM, but overall effectiveness is determined by the extent to which providers comply with patient preferences.* ***Key words:*** *preferences and quality of life; clinical prediction rules; risk stratification; decision aids; patient decision making; provider decision making; physician-patient communication; shared decision making.* ***(Med Decis Making 2011;31:93–107)***

Received 14 July 2009 from the Department of Medicine, Boston University School of Medicine, Boston, MA (PCS, JTG, PAR, MAL, SM, PD); Medical Oncology, Dana Farber Cancer Institute, Boston, MA (KE); Decision Research, Eugene, OR (EP); Data Coordinating Center, Boston University School of Public Health, Boston, MA (SE, CC); Department of Medicine, University of North Carolina, Chapel Hill, NC (MP); Department of Epidemiology, Boston University School of Public Health, Boston, MA (MP); and Department of Biostatistics, Boston University School of Public Health, Boston, MA (TCH). Portions of this paper were presented at the 2008 Annual Meeting of the Society for Medical Decision Making, Philadelphia, PA. Revision accepted for publication 27 January 2010.

Address correspondence to Paul C. Schroy III, MD, MPH, Boston Medical Center, 85 East Concord Street, Suite 7715, Boston, MA 02118; telephone: (617) 638-8335; fax: (617) 638-6529; e-mail: paul.schroy@bmc.org.

DOI: 10.1177/0272989X10369007

Colorectal cancer (CRC) is the second leading cause of cancer-related death and the third most commonly diagnosed cancer among men and women in the United States. Screening has been shown to be a cost-effective strategy for reducing both CRC mortality and incidence[1] and is now widely endorsed by most authoritative groups.[2,3] Despite this endorsement, however, screening rates remain relatively low, partly due to poor patient acceptance and adherence. Data from the 2005 National Health Interview Survey suggest that while screening rates have improved, approximately 50% of eligible Americans have not had appropriate screening.[4]

Shared decision making (SDM) has been advocated as a potentially effective strategy for

increasing patient acceptance and adherence to CRC screening recommendations.[2,3] CRC screening is ideally suited for this approach given the availability of multiple strategies with distinct advantages and disadvantages, the lack of consensus regarding an optimal strategy, and historical ineffectiveness of the more traditional paternalistic approach in which providers assume full responsibility for the decision-making process. Further support is derived from studies that find that patients have distinct preferences for the different screening strategies,[5–18] that providers often misperceive patient preferences,[9] and that many patients support the SDM approach for CRC screening.[16]

SDM is an interactive process in which patients and their health care providers form a partnership to exchange information, clarify values, and negotiate a mutually agreeable medical decision.[19,20] Unfortunately, SDM has been difficult to implement into routine clinical practice in part due to lack of time, lack of clinician expertise, and lack of resources.[21] To circumvent several of these barriers, decision aids have been developed to facilitate informed decision making (IDM) outside of the clinical encounter.[22] IDM is a process in which patients receive sufficient information about the risks, benefits, limitations, alternatives, and uncertainties of a clinical condition or disease to make a value-concordant decision and participate in the decision-making process at a desired level.[23,24] Thus, it is not as collaborative with the provider as SDM but still includes a key element of patient empowerment. As with any decision aid, decision aids for CRC screening should provide, at a minimum, sufficient information about the pros and cons of the recommended options to enable users to identify a value-concordant preferred option.[25] Besides facilitating IDM, decision aids also have the potential to facilitate SDM by improving the quality and efficiency of the patient-provider encounter and by empowering users to participate in the decision-making process.[25] Alternatively, enabling patients to identify a preferred screening option outside of the clinical encounter could have a detrimental effect on the decision-making process in situations where patient and provider preferences differ by inducing decisional conflict and/or dissatisfaction with the provider recommendation.

Studies to date have clearly demonstrated that existing decision aids for CRC screening enable users to identify a preferred screening option,[6–9,12,13,15–18,26] reduce decisional conflict,[26] and increase interest in screening.[8,10,13,15] The extent to which decision aids facilitate effective SDM and increase adherence, which are of critical importance to the utility of such tools in clinical practice, however, is less well established. To address the former shortcoming, we report herein the interim results of a randomized controlled clinical trial aimed partly at evaluating the impact of a novel, interactive computer-based decision aid on relevant measures of SDM, including patient knowledge, patient preferences, satisfaction with the decision-making process, concordance between patient preference and test ordered, and screening intentions.

METHODS

Decision Aid Development, Format, and Usability Testing

Development of our decision aid was guided by constructs of the Ottawa Decision Support Framework.[27,28] The actual content of the tool was derived from systematic reviews of available evidence on the cost-effectiveness and attributes of the different screening strategies,[1,29,30] a systematic review of existing decision aids,[25] and expert opinion. We also conducted a series of focus groups of racially/ethnically diverse previously screened men (n = 7), previously screened women (n = 10), unscreened men (n = 5), and unscreened women (n = 9) recruited from a convenience sample of primary care patients seen at the target sites to determine key factors (e.g., knowledge, attitudes, beliefs, literacy, numeracy, etc.) that needed to be incorporated in the tool.

The prototype version of the decision aid was built using Web-based technology but initially formatted for DVD use on portable computer stations to provide maximum flexibility for use in the ambulatory care setting during the clinical trial. The decision aid uses an audiovisual and touch-screen format to simplify use for individuals with limited literacy and/or computer skills. It is comprised of a series of modules in which professional actors playing the role of a black, Hispanic female moderator and a white, non-Hispanic male physician convey relevant information via on-screen video, animation, and/or graphics. The modules (Figure 1) include 1) an introductory segment that briefly discusses the importance of screening, intended purpose of the tool, and instructions in its use; 2) an overview of the epidemiology of CRC, natural history, rationale for screening, benefits of screening, the availability of multiple screening options, and

the lack of consensus regarding a best screening method; 3) brief descriptions of the 5 recommended screening methods (fecal occult blood testing [FOBT], flexible sigmoidoscopy, the combination of FOBT plus flexible sigmoidoscopy, double-contrast barium enema, and colonoscopy) endorsed by the American Cancer Society, US Preventive Services Task Force, and GI Multi-Society Task Force on Colorectal Cancer at the time[2,3]; 4) audio and visual (i.e., traffic light graphic) comparisons of each screening method with respect to individual test-features including accuracy, inconvenience, discomfort, recommended frequency of testing, complications, and need (or lack thereof) for further diagnostic studies if the screening test result is positive; 5) a summary of the different test features for each screening strategy with optional links to more detailed information about the preparation or test itself, as well as vignettes from a racially diverse group of patients describing their experience with a particular test; 6) a decision-making module in which users are asked to identify a screening preference or lack thereof, using the discrete choice method, rank order test features influencing their selection, and assess whether out-of-pocket payments (if not covered) would alter their choice; and 7) a concluding segment in which the narrator encourages the user to discuss screening and their screening preferences with their doctor, acknowledging that the best CRC screening test is the test that gets done. Users can navigate forward or backward through the tool using either the touch screen or a mouse; they also have the option of repeating segments and printing out key information. Once developed, the tool was reviewed by the Massachusetts Department of Public Health's Colorectal Cancer Working Group for content validity and cultural sensitivity.

A modified version of the decision aid was also created, which includes the Web-based ''Your Disease Risk'' (YDR) CRC risk assessment tool (http://www.yourdiseaserisk.wustl.edu). The risk assessment tool was placed just after the introductory segment. The intent of the tool was to assess the extent to which personalized risk estimates influenced decision making. Based on the available literature at the time,[31] we postulated that personalized risk information might positively influence adherence among patients deemed to be at above-average risk.

Usability testing was performed prior to its implementation, in accordance with recommendations by Nielsen,[32,33] to assess ease of learning, efficiency of use, and user satisfaction. Based on observational data and feedback from 2 rounds of testing with 5 different users per test, the prototype was revised to enhance functionality and deemed ready for clinical use without further testing.

Study Design

A randomized controlled trial was initiated in April 2005 to evaluate the impact of our decision aid on SDM and patient adherence to CRC screening recommendations. Eligible patients were instructed to arrive 1 hour before a prearranged office visit with their primary care provider. After obtaining informed consent, patients were administered a pretest comprised of 28 close-ended questions that assessed knowledge, beliefs, attitudes, and behaviors related to CRC screening, as well as level of desire for participating in decision making related to CRC screening.[34] The pretest was administered by 1 of 3 trained research assistants in a private office located in one of the ambulatory care clinics at Boston Medical Center or the South Boston Community Health Center. After completing the pretest, patients were randomized to 1 of 2 intervention arms (decision aid plus YDR personalized risk assessment tool with feedback or decision aid alone) or a control arm after stratification by provider. Patients randomized to the control arm reviewed a modified version of ''9 Ways to Stay Healthy and Prevent Disease,'' previously posted on the Harvard Center for Cancer Prevention Web site, which discussed generic lifestyle changes other than screening for minimizing risk of preventable diseases. After completing the interactive computer session, patients then met with their providers to discuss screening and identify a preferred screening strategy. Providers received written notification in the form of a flyer hand delivered by the patient acknowledging that the patient was participating in the ''CRC decision aid'' study at the time of the visit to ensure that screening was discussed; no information was provided about preferences or factors influencing choice for patients in the intervention arm. Before leaving the clinic, subjects were administered a posttest, which assessed whether CRC screening was discussed, whether a screening strategy was chosen, patient satisfaction with the decision-making process, and screening intentions; the posttest also reassessed knowledge, beliefs, and attitudes related to CRC screening.

Setting

The study was conducted at 2 urban ambulatory care sites. The first, Boston Medical Center, is

Figure 1 Decision aid. Representative screens from the different segments of the tool including the introductory module (A); overview of colorectal cancer and colorectal screening (B); brief descriptions of each screening option (C); list of test features discussed (D); comparisons of screening options with respect to individual test features (E); summaries of attributes for each option with links to more detailed information about the preparation, procedure itself, and patient testimonials (F); and the decision-making module, where users are asked to identify a preferred option (G) and rank order test features influencing choice (H). One version of the tool also includes the ''Your Disease Risk'' risk assessment tool, which calculates personalized 10-year estimates for developing colorectal cancer (I).

a private, nonprofit academic medical center affiliated with the Boston University School of Medicine, which serves a mostly minority patient population (only 28% white, non-Hispanic). The second, the South Boston Community Health Center, is a community health center affiliated with Boston Medical Center, which serves a mostly white, non-Hispanic, low-income patient population. Both sites use the same electronic medical record system (Logician Centricity™, General Electric Company). The study protocols were approved by the Boston University Medical Campus Institutional Review Board, which was responsible for overseeing human studies research at both participating institutions.

Study Population and Recruitment Process

The study sample was comprised of average-risk patients under the care of one of the primary care providers at Boston Medical Center or the South Boston Community Health Center. Patients were deemed eligible if they were 50 to 75 years of age and had no

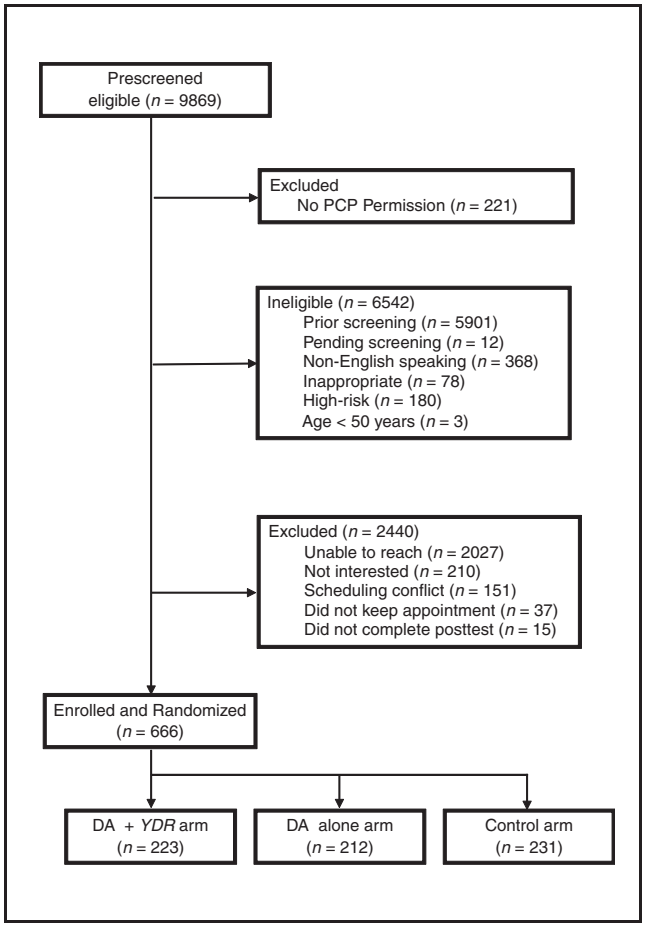


Figure 2 Study flow diagram. DA = decision aid; YDR = ''Your Disease Risk'' risk assessment tool.

prior structural CRC screening examinations. Potential subjects meeting any of the following criteria were excluded: 1) prior CRC screening by any method other than FOBT; 2) high-risk condition (personal history of colorectal cancer or polyps, family history of colorectal cancer or polyps involving one or more first-degree relatives, or chronic inflammatory bowel disease); 3) lack of fluency in written and spoken English; or 4) comorbidities that preclude CRC screening by any recommended method.

Three different recruitment strategies were used during the course of the study. The vast majority of patients ($N = 637$) were recruited using an investigator-initiated ''opt-out'' approach in which patients due for screening were identified from monthly audits of Boston Medical Center's electronic medical record 2 to 4 weeks prior to a scheduled office visit and contacted directly by telephone by one of the research assistants if deemed appropriate by the patient's primary care provider. Those expressing interest were provided with a brief overview of the study, evaluated for eligibility, and invited to participate using a passive informed consent process. Two other strategies, including a provider-initiated, ''opt-in'' electronic flagging approach ($N = 12$) and a provider-mediated, ''out-in'' letter approach ($N = 17$), were used initially but discontinued after 6 months due to low enrollment. As opposed to the ''opt-out'' investigator approach whereby potential subjects could decline to be recruited after being contacted by the research team without prior permission, the ''opt-in'' approach required that potential subjects grant permission to be contacted from the onset, and a nonresponse prohibited further communication. The relative cost-effectiveness of each of our approaches has recently been published.[35]

Provider Characteristics and Training

Fifty full-time primary care providers, including 47 board-certified general internists and 3 nurse practitioners, practicing at both Boston Medical Center and the South Boston Community Health Center participated in the study. A pretrial survey of 30 participating providers indicated that virtually all (97%) preferred a SDM approach when selecting an appropriate screening strategy for their patients. Pretrial seminars were conducted at both sites to educate providers about the current status of CRC screening, highlighting the recommendation for SDM, to provide an overview of the study design, and to elicit support. Because of provider turnover, brief annual meetings were also conducted to ensure that new providers were aware of the study, understood its design, and expressed a willingness to participate. The meetings also provided a venue for informing participating providers about the status of recruitment and addressing any logistical problems that they were experiencing related to the study. No formal training in SDM was undertaken.

Outcome Measures

The key outcome measures of interest for assessing the impact of our decision aid on SDM were patient knowledge, patient preferences, satisfaction with the decision-making process, screening

intentions, and test concordance (i.e., agreement between patient preference and test ordered).

Knowledge was assessed at baseline (pretest) and at the time of the exit survey (posttest) based on responses to a 12-item questionnaire (True/False/Don't know) that inquired about CRC risk factors, the rationale and goals of screening, and age at which screening should begin (see Appendix 1 for individual questions). The content was derived from key messages endorsed by the National Colorectal Cancer Roundtable[36] and the Massachusetts Department of Public Health[37] for public education. Cumulative knowledge scores (range, 0–12) were derived by summing correct responses to the 12 individual knowledge questions.

Patient preferences and factors influencing choice of options were obtained from response data captured electronically in the decision aid's decision-making module.

Patient satisfaction with the decision-making process was assessed on the posttest using the validated 12-item Satisfaction with the Decision-Making Process Scale (Appendix 2),[38] which has excellent reliability (Cronbach $\alpha = 0.85$) and discriminant validity. Five ordered response categories were used for each item. Each response was assigned a point score ranging from 1 for ''strongly disagree'' (or ''poor'') to 5 for ''strongly agree'' (or ''excellent''). A cumulative score was calculated based on the summed response scores for each item (maximum score = 60). Mean item substitution was used to impute missing data for patients answering between 8 and 11 items; patients answering fewer than 8 items were excluded from analysis.

Screening intentions were also assessed as part of the posttest. Subjects were asked how sure they were that they would schedule an appointment to get screened for CRC and how sure they were that they would complete the screening test they scheduled. An ordered 5-point response frame was used ranging from ''not at all sure'' to ''completely sure.''

Concordance between patient preferences and test ordered was assessed for the 2 intervention groups only because preferences were not elicited from the control group. Test ordered was ascertained from the ''Orders'' section of the electronic medical record. Concordance was defined as the percentage of patients who had their preferred screening test ordered.

Statistical Analyses

Based on crude estimates of baseline adherence, we calculated that a target sample of 825 patients provided 80% power of detecting a 54% versus 40% pair-wise difference in the percentage of patients completing a CRC screening, using a Bonferroni adjustment to overall α level of 0.05 to account for the 3-group study design.

As a check on randomization, the 3 study groups were compared on demographic characteristics, prior screening, and desired role in decision making through the χ^2 test of independence. The 3 study groups were also compared on cumulative pretest and posttest knowledge scores, satisfaction with the decision-making process (SDMP) scores, and intention scores through separate 1-factor analysis of covariance (ANCOVA); Bonferroni-adjusted multiple comparison procedure was used to investigate pair-wise differences following a significant 3-group comparison. ANCOVA was also used for subgroup analyses comparing SDMP scores across the 3 study groups after stratification by desired role in decision making (mostly patient, shared, and mostly provider). For the SDMP analyses, mean item substitution was used to impute missing data for patients who answered at least 8 items. To evaluate the possible bias of this approach, we employed multiple imputation analysis to generate 5 data sets using the expectation-maximization (EM) algorithm and found that the results agreed closely with those derived using mean item substitution. Differences between study groups on adjusted means from the ANCOVA are described through effect sizes (d) calculated as the difference in adjusted means divided by the pooled standard deviation estimate from the ANCOVA. Descriptive statistics were used to describe patient preferences for the 5 screening test options and factors influencing choice; associations between demographic characteristics and patient preferences (colonoscopy v. FOBT) were examined through a series of multiple logistic regressions. Test concordance, that is, the association between patient preference and test ordered, was examined through the χ^2 test of independence. Associations between test concordance and desired role in decision making were tested through the χ^2 test of independence, whereas associations between test concordance and both SDMP and intention scores were compared through the independent sample t tests. We controlled for study site (Boston Medical Center and South Boston Community Health Center) in the analyses of all outcome measures, using ANCOVA regression models for measurement outcomes and Cochran-Mantel-Haenszel analyses or logistic regression analyses for categorical outcomes. Data are expressed as mean [standard deviation], unless otherwise stated. P values less than 0.05 were deemed significant.

Table 1 Characteristics of Study Participants ($N = 666$)

Characteristic	DA + YCR ($n = 223$)	DA Alone ($n = 212$)	Control ($n = 231$)	P Value
Age, *n* (%)				
<65 years	181 (81)	182 (86)	191 (83)	0.41
≥65 years	42 (19)	30 (14)	40 (17)	
Sex, *n* (%)				0.79
Female	137 (61)	125 (59)	135 (58)	
Male	86 (39)	87 (41)	96 (42)	
Ethnicity, *n* (%)				0.75
Non-Hispanic	209 (94)	202 (95)	217 (94)	
Hispanic	14 (6)	10 (5)	14 (6)	
Race, *n* (%)				
Black	139 (62)	124 (58)	155 (67)	0.21
White	78 (35)	76 (36)	71 (30)	
Asian	2 (1)	6 (3)	1 (1)	
Other	4 (2)	6 (3)	4 (2)	
Education, *n* (%)				0.56
≥ High school	170(76)	159 (76)	165 (73)	
< High school	52 (24)	49 (24)	62 (27)	
Insurance, *n* (%)				
Private/HMO	78 (39)	73 (39)	77 (36)	0.61
Medicare	65 (33)	47 (25)	66 (31)	
Medicaid	44 (22)	54 (29)	53 (25)	
Free care	9 (4)	11 (6)	19 (5)	
None	3 (2)	2 (1)	6 (3)	
Prior FOBT				
Yes	29 (13)	30 (14)	34 (15)	0.90
No	189 (87)	180 (86)	196 (85)	
Desired role in decision making, *n* (%)				
Mostly patient	60 (27)	53 (25)	70 (30)	0.68
Shared	120 (54)	120 (57)	115 (50)	
Mostly doctor	43 (19)	39 (18)	46 (20)	

Note: DA = decision aid; YDR = ''Your Disease Risk''; HMO = health maintenance organization; FOBT = fecal occult blood testing.

RESULTS

Patient Characteristics

Of the 9648 patients identified as potentially eligible for screening (aged 50–75 years) for whom permission to contact was granted, 6542 (68%) were deemed ineligible (mostly due to prior screening), and 2440 (25%) were excluded. Reasons for exclusion were inability to contact ($n = 2027$), disinterest ($n = 210$), scheduling conflict ($n = 151$), failure to keep appointment ($n = 37$), or failure to complete posttest ($n = 15$) (Figure 2). The remaining 666 patients (62% of eligible subjects contacted) were enrolled and randomized to the decision aid plus ''Your Disease Risk'' (YDR) arm ($n = 223$), decision aid alone arm ($n = 212$), or control arm ($n = 231$).

As shown in Table 1, the 3 study arms were well balanced with respect to patient age, sex, race, ethnicity, education, prior FOBT, insurance status, and decision-making preference. Overall, the study group was predominantly less than 65 years of age, female, non-Hispanic, and black with at least a high school degree. Although most had some form of health care insurance, nearly two thirds were covered by Medicare, Medicaid, or the Massachusetts' ''free care'' program. Most had no prior experience with FOBT. Importantly, the majority preferred a patient-dominant (28%) or SDM approach (53%) for selecting a preferred CRC screening option.

Knowledge

Mean [standard deviation] cumulative pretest knowledge scores were comparable ($P = 0.91$) for the 3 groups (decision aid plus YDR, 7.6 [2.8]; decision aid alone, 7.7 [2.9]; control, 7.5 [2.7]). Cumulative posttest scores, however, were significantly higher

Table 2 Patient Preferences and Most Important Test Feature Influencing Test Preference (Intervention Groups Only, $N = 435$[a])

	Patient Preference, *n* (%)					
	Colon	**FOBT**	**Flex Sig**	**FOBT + Flex Sig**	**DCBE**	**None**
Study group[a]						
DA + YDR	132 (60)	53 (24)	13 (6)	6 (2)	8 (4)	8 (4)
DA alone	120 (57)	58 (28)	11 (5)	5 (3)	9 (4)	7 (3)
Combined[b]	252 (59)	111 (26)	24 (5)	11 (3)	17 (4)	15 (3)
Most important test feature[c]						
Accuracy	205 (81)	16 (14)	11 (46)	6 (54)	4 (24)	—
Preparation	13 (5)	20 (18)	3 (12)	1 (9)	2 (12)	—
Amount of discomfort	3 (1)	34 (31)	2 (9)	2 (18)	2 (12)	—
Inconvenience	4 (2)	25 (23)	4 (17)	0	4 (24)	—
Risk of complications	5 (2)	9 (8)	3 (12)	2 (19)	4 (24)	—
Frequency	14 (6)	0	0	0	1 (6)	—
Need for further testing if results abnormal	8 (3)	1 (3)	1(4)	0	0	—

Note: DA = decision aid; YDR = ''Your Disease Risk''; Colon = colonoscopy; FOBT = fecal occult blood testing; Flex Sig = flexible sigmoidoscopy; DCBE = double-contrast barium enema.
a. Percentages relate to rows.
b. Data missing, $n = 5$.
c. Percentages relate to columns.

($P < 0.001$) for the 2 intervention groups (decision aid plus YDR, 10.7 [1.8]; decision aid alone, 10.9 [1.6]) compared with the control group (8.6 [2.7]), with differences corresponding to large effect sizes of $d = 1.15$ and $d = 1.27$ for the decision aid plus YDR group and decision aid alone group versus control, respectively. The mean increase in scores from pretest to posttest was also significant ($P < 0.001$) for both intervention groups (decision aid plus YDR, 3.0 [2.5]; decision aid alone, 3.2 [2.6]) but not the control group (0.8 [2.2]). There were no significant differences in cumulative posttest scores or change in scores between the 2 intervention groups.

Patient Preferences and Factors Influencing Choice

Screening test preferences for patients randomized to the 2 intervention arms are depicted in Table 2. Colonoscopy was preferred by a majority of patients (59%), followed by FOBT (26%), flexible sigmoidoscopy (5%), double-contrast barium enema (4%), and FOBT plus flexible sigmoidoscopy (3%); only 3% either declined testing or could not identify a preferred option. No significant differences were observed between the 2 intervention groups. Patients who chose colonoscopy were most likely to identify test accuracy (81%) as the most important feature influencing their choice; in contrast, patients who chose FOBT were most likely to identify concerns about discomfort (31%), inconvenience (23%), and bowel preparation (18%) as the most influential features. Logistic regression analysis found no significant independent associations between demographic factors (age, sex, race, ethnicity, education, insurance status, or study site) and preference for colonoscopy versus FOBT (data not shown).

Satisfaction with the Decision-Making Process (SDMP)

Overall, 636 subjects (96% overall) responded to ≥ 8 items of the SDMP scale and were included in the analysis; there were no significant differences in the percentage of patients answering all 12 items versus 8 to 11 items versus < 8 items (excluded) across the 3 study groups ($P = 0.31$). As shown in Table 3, mean SDMP scores were significantly higher for the 2 intervention arms compared to controls, with differences corresponding to moderate effect sizes of 0.53 and 0.61 for the decision aid plus YDR and decision aid alone groups versus control, respectively. Scores for the 2 intervention groups were comparable. Subgroup analysis found that satisfaction was also higher among intervention patients who preferred a shared or patient-dominant role in decision making; a similar trend was observed for patients who preferred a provider-dominant approach, but the differences between the intervention and control groups did not achieve statistical significance.

Table 3 Satisfaction with the Decision-Making Process (SDMP) Scores

	DA + YDR	DA Alone	Control	*P* Value
Overall	50.5 (6.2) *n* = 214	50.7 (6.2) *n* = 205	46.7 (7.9) (*n* =217)	<0.001[a]
Decision-making preference				
Mostly patient	49.4 (6.2) *n* = 58	50.2 (6.5) *n* = 50	46.0 (8.3) *n* = 66	0.01[a]
Shared	50.8 (6.4) *n* = 115	50.6 (6.3) *n* = 116	46.6 (7.9) *n* = 108	<0.001[a]
Mostly provider	51.4 (5.4) *n* = 41	51.5 (5.3) *n* = 39	49.0 (6.7) *n* = 43	0.06

Note: DA = decision aid; YDR = ''Your Disease Risk''. Data expressed as mean (standard deviation); maximum score = 60.
a. One-factor ANCOVA showed a significant difference in satisfaction between the 3 study groups, with pair-wise comparisons showing no significant differences between the 2 intervention groups and significantly lower satisfaction for those in the control group overall and after stratification by decision-making preference.

Intentions

Mean intention scores [standard deviation] were significantly higher ($P < 0.001$) for the 2 intervention groups (decision aid plus YDR, 4.3 [1.0]; decision aid alone, 4.4 [1.0]) compared to the control group (3.9 [1.4]) when asked ''How sure are you that you will schedule a colorectal cancer screening test?'' Mean intention scores were also significantly higher ($P < 0.001$) for the 2 intervention groups (both 4.3 [1.0]) compared to controls (3.9 [1.3]) when asked ''How sure are you that you will complete a colorectal cancer screening test?'' Differences in intention to schedule or complete a screening test for the 2 intervention groups versus control correspond to moderate effect sizes ranging between 0.36 and 0.44. Scores were comparable for the 2 intervention groups.

Concordance between Patient Preference and Test Ordered

Concordance between preference and test ordered is shown in Table 4. Among the 415 patients expressing a preference, 244 (59%) had their preferred test ordered, 79 (19%) had an alternate test ordered (colonoscopy, 85%; FOBT, 14%; flexible sigmoidoscopy, 1%), and 92 (22%) had no test ordered. For individual tests, concordance between patient preference and test ordered varied from 79% for colonoscopy to $\leq$30% for all other options. For the discordant group, virtually everyone (96%) who preferred a test other than colonoscopy had a colonoscopy ordered. Patients who preferred colonoscopy were more likely to have any test ordered than those who preferred a test other than colonoscopy (83% v. 70%; $P < 0.002$).

As shown in Table 5, there was no association between concordance and desired role in decision making, demonstrating that patients who preferred a patient-dominant or SDM style were no more likely to have a concordant test ordered than those who preferred a provider-dominant style. There was also no significant association between SDMP scores and concordance. Importantly, however, intention scores were significantly higher when there was concordance between patient preference and test ordered compared to when they differed. The positive associations between test concordance and screening intentions replicated significantly within the independent test preference groups (colonoscopy and ''other'').

Overall, patients in the intervention arms were more likely to have a test ordered than patients in the control arm (75% v. 68%; $P < 0.05$), regardless of whether there was concordance or discordance with test preference (data not shown). Colonoscopy (82%) was the most commonly ordered test for control patients with any test ordered, followed by FOBT (12%) and flexible sigmoidoscopy (1%), thus affirming a strong provider preference for colonoscopy.

DISCUSSION

Most authoritative groups, including the US Preventive Services Task Force, American Cancer Society, US Multi-Society Task Force on Colorectal Cancer, and American College of Radiology,[2,3] endorse a SDM approach when selecting an

Table 4 Concordance between Patient Preference and Test Ordered (Intervention Groups Only[a])

Test Ordered	Patient Preference, *n* (%)					
	Colon (*n* = 252)	FOBT (*n* = 111)	Flex Sig (*n* = 24)	DCBE (*n* = 17)	FOBT + Flex Sig (*n* = 11)	Overall (*n* = 415)
Same	199 (79)	34 (30)	5 (21)	4 (24)	2 (18)[b]	244 (59)
Different	10 (4)	44 (40)	12 (50)	7 (41)	6 (55)	79 (19)
None	43 (17)	33 (30)	7 (29)	6 (35)	3 (27)	92 (22)

Note: Colon = colonoscopy; FOBT = fecal occult blood testing; Flex Sig = flexible sigmoidoscopy; DCBE = double-contrast barium enema.
a. Intervention patients with no preference (*n* = 15) or missing data (*n* = 5) not included.
b. ''Same'' infers that at least 1 of the 2 preferred tests was ordered.

appropriate CRC screening strategy, but implementation into clinical practice encounters many barriers.[21] The purpose of this study was to assess the extent to which use of a novel CRC screening decision aid circumvents some of these barriers and facilitates effective SDM in the primary care setting. Like similar such studies,[6–9,12,13,15–18,26] our study finds that the tool enables users to identify a preferred screening option based on the relative value they place on individual test features. Our study also finds that while users are more knowledgeable, more satisfied with the decision-making process, and more intent to undergo screening than nonusers overall, screening intentions and test ordering are negatively influenced in situations where patient and provider preferences differ, regardless of a patient's desired role in decision making. Together, these observations suggest that the utility of our decision aid for promoting effective SDM is dependent upon the extent to which providers are willing to comply with an informed patient's screening preferences. Importantly, however, screening intentions were still higher for discordant users than for nonusers, suggesting that use of the decision aid is nonetheless superior to nonuse.

A critical question that arises when interpreting these findings relates to whether our outcomes of interest are valid measures of effective SDM. As previously noted, SDM has been defined as a process in which both the patient and provider engage in information exchange, deliberation (value clarification), negotiation, and mutual decision making.[19,20] Accordingly, we opted to include patient knowledge, SDMP, concordance between patient preference and test ordered, and screening intentions as appropriate measures of SDM within the context of this conceptual framework.

The notion of information exchange implies that there was a reciprocal exchange of information between the patient and health care provider. Because the primary intent of our decision aid was to educate users about the rationale for screening and the pros and cons of the different screening options so as to make an informed choice about screening and test preference, our knowledge outcome is a more reliable measure of IDM than SDM. The finding that knowledge scores did not increase after the provider encounter for the control group, however, suggests that most providers failed to convey relevant information about the rationale for screening, risk factors, goals, and benefits when discussing CRC screening with their patients.

The SDMP outcome provides a useful measure of patient involvement in the decision-making process from the patient perspective. As with decision aids for other conditions,[25] our findings suggest that our tool also empowered users to participate in the decision-making process at a desired level, particularly those who preferred a shared or patient-dominant role in decision making. We speculate that our tool's format and inclusion of specific messages encouraging patients to take a proactive role in the decision-making process provided meaningful guidance in deliberation and/or communication.

Our concordance and intention outcomes are arguably the most relevant measures of ''effective'' SDM. Concordance provides a measure of the extent to which providers are willing to respect patient autonomy in the decision-making process and comply with patient preferences when selecting an appropriate screening test. Intention provides a measure of the degree to which the patient is committed to the chosen course of action and, in the case of SDM, provides perspective on the extent to which the final decision reflects successful negotiation and mutual agreement. Intention is also a powerful predictor of CRC screening adherence.[39,40] In this study, we found that concordance was high for colonoscopy but relatively low for the other options, suggesting that providers were much more likely to

Table 5 Association between Test Concordance and Desired Role in Decision-Making Preference, Satisfaction with Decision-Making Process, and Screening Intentions (Intervention Groups Only)

	Test Concordance[a]		
Outcome	**Same**	**Different**	***P* Value**
Desired role in decision making, *n* (%)			0.9[d]
Mostly patient	60 (75)	20 (25)	
Shared	138 (75)	45 (25)	
Mostly doctor	45 (75)	15 (25)	
SDMP score[b]	51.3 (6.4)	49.4 (5.8)	0.20[e]
	n = 241	*n* = 75	
Intention[c]			
Schedule test	4.6 (0.7)	4.2 (1.1)	<0.001[e]
	n = 241	*n* = 75	
Complete test	4.6 (0.7)	4.2 (1.1)	< 0.001[e]
	n = 243	*n* = 74	

Note: SDMP = satisfaction with the decision-making process.
a. Agreement between patient preference and test ordered.
b. Mean cumulative scores (standard deviation); maximum = 60.
c. Mean scores standard deviation, where 5 = ''very sure'' and 1 = ''very unsure.''
d. Cochran-Mantel-Haenszel χ^2 analyses controlling for site.
e. ANCOVA controlling for site.

comply with patient preferences that agreed with their own screening preferences. We also found that the likelihood of having the preferred test ordered was unaffected by the extent to which patients wished to engage in the decision-making process, suggesting that most providers either fail to assess their patients' desired level of participation or assume decisional authority, regardless of the preferred role. Perhaps most importantly, we found that when there was concordance between patient and provider preferences, patients had stronger screening intentions and a greater likelihood of having any screening test ordered compared to when patient and provider preferences differed. In the aggregate, these observations suggest that our decision aid facilitates effective SDM in settings where providers truly endorse a SDM approach and are willing to comply with patient preferences when selecting an appropriate screening test. In settings where providers feel compelled to endorse a single screening option, regardless of patients' desired level of participation in the decision-making process, use of the tool may compromise effectiveness, possibly due to enhanced decisional conflict and/or dissatisfaction with the process or decision itself.

One of the principal goals of both IDM and SDM for CRC screening is the elicitation of a value-concordant patient preference for one of the recommended screening tests. Our findings corroborate a body of evidence demonstrating that patients have distinct CRC screening test preferences that are influenced by the value they place on one or more test features.[5–18] As in our prior study,[16] we found that colonoscopy was preferred by a majority (59%) of intervention patients, most of whom identified test accuracy as the predominant reason for their selection; in contrast, most of the remaining patients (26%) preferred FOBT primarily because of concerns about discomfort, inconvenience, and bowel preparation. Both Hawley et al.[18] and Marshall et al.[17] also observed a dominant preference for colonoscopy compared to FOBT, flexible sigmoidoscopy, and barium enema. Although the surge in media campaigns promoting colonoscopy in recent years might explain this predilection for colonoscopy, a similar trend has not been reported in other recently published studies.[12–15] Differences in the study population, options discussed, framing, and methodologies used to elicit preferences may be important factors in accounting for these disparate results. Collectively, however, these findings support the rationale for eliciting patient preferences within the context of IDM and/or SDM.

Our study also provides new evidence suggesting that providers also have a dominant preference for colonoscopy. Using test ordered as a proxy for provider preference, we found that providers referred the majority (~82%) of control patients for colonoscopy and that providers were unlikely to order tests other than colonoscopy, regardless of patient preferences, for the 2 intervention groups. This predilection for colonoscopy mirrors national trends in

utilization[4] and may be explained by highly publicized studies affirming superior accuracy for detecting advanced neoplasia compared to other screening modalities,[41–44] expanded coverage by Medicare and many health insurers, and an increase in patient demand due to heightened public awareness efforts.[45] The extent to which liability concerns, sociodemographic factors, health care coverage, patient comorbidities, and/or lack of adequate systems for follow-up (as in the case of FOBT) influence provider recommendations when dealing with individual patients is less well defined. A reliable risk prediction model that could accurately stratify patients into different risk categories for the presence of advanced colorectal adenomas and cancer at screening colonoscopy would be invaluable in this regard because it would bestow providers with objective decisional support when considering patient preferences and thereby enhance the effectiveness of SDM. In the interim, however, the growing tendency to promote colonoscopy independent of patient preferences undermines the spirit of SDM and its potential effectiveness as a strategy for increasing CRC screening rates.

Our decision aid has several features that distinguish it from other currently available CRC screening decision aids. First, the tool employs culturally sensitive video narratives, simple graphics and animation to enhance its appropriateness, and acceptability and comprehensibility for use by a diverse target audience with variable literacy skills. Second, it provides an overview of CRC and CRC screening that incorporates key messages endorsed by the National Colorectal Cancer Roundtable to heighten awareness and entice interest in screening.[36] Third, one version also incorporates the YDR risk assessment tool to enable users to incorporate personalized risk estimates into their decision making. Of note, however, we observed no significant associations between personalized risk feedback and patient preferences or other outcomes of interest. We speculate that the lack of difference between the 2 intervention arms may be due to the offsetting effects of informing some patients in the YDR arm that they were at above-average risk and others below average risk. Alternatively, the lack of difference may also be due to the fact that the information was disregarded by many of those receiving personalized feedback, possibly because it was perceived to be incorrect.[46] Fourth, unlike more complex decision-making approaches, such as conjoint analysis[18] and analytic hierarchy analysis,[12,26] the tool employs descriptive attribute-based comparisons of the different screening tests, as well as option-based summaries, to enable users to identify a value-concordant preference. Fifth, it provides optional links to more detailed information about the preparation, test itself, and patient testimonials to ensure that users have as much information as desired to make an informed choice. Lastly, it provides users with a no-screening option and explores reasons for their decision. The major drawbacks of the tool in its current form are its length, if users opt to view all segments, and its linear arrangement. Enhancements that permit tailored navigation to fit the informational needs of users are required for optimal Internet dissemination.

Our study has several notable strengths. First, the study provides new insight into the utility of decision aids for facilitating SDM related to CRC screening by assessing the extent to which such tools enhance satisfaction with the decision-making process and the extent to which concordance between patient preferences and provider preferences influence screening intentions and test ordering. Second, the use of a randomized controlled study design, a large sample size, and a racially diverse study population enhances both the internal and external validity of its findings. Third, implementation of the decision aid as a point-of-contact intervention attests to the feasibility of use in the primary care setting. Finally, the inclusion of mostly unscreened patients minimizes potential confounding.

Our study also has several notable limitations. First, the lack of provider blinding might have negatively influenced the magnitude of the interventions' effect on several outcomes of interest, most notably satisfaction with the decision-making process. Second, no attempt was made to assess the quality of the patient-provider discussion. Even though SDMP was universally high (albeit higher in the intervention groups), recent data suggest that most patient-provider discussions related to CRC screening fail to incorporate key elements of IDM.[47–49] Third, no attempt was made to assess factors influencing provider preferences, as previously noted, nor provider satisfaction related to either the use of the decision aid or the decision-making process. Fourth, we failed to explore whether users experienced greater decisional conflict or uncertainty when their preferences differ from those of their provider. Fifth, we opted to exclude cost information in our tool when discussing attributes because of conflicting data regarding its impact on decision making in this setting[7,50,51] and because of

the complexity of presenting such information in a meaningful way given the variability in coverage under different health plans. Lastly, despite its obvious importance, we opted to exclude data on patient adherence in this interim analysis because of incomplete follow-up. Although extrapolation of our findings suggests that failure to comply with patient preferences may have negative consequences on screening behavior, it could be that providers who can effectively communicate their reasoning for recommending colonoscopy and facilitate its completion may be successful in getting their patients screened, regardless of their preferences. Existing data, however, would suggest that providers who focus on colonoscopy tend to adopt a more paternalistic approach with little exchange of information or attention to barriers.[52]

In conclusion, our study finds that our decision aid facilitates SDM related to CRC screening by enabling patients to identify a preferred screening option based on the values they place on individual test features and screening intentions. Although use of the tool also enhances SDMP, this empowerment may compromise the effectiveness of SDM in situations where patient and provider preferences differ. Future studies are needed to better understand factors that influence provider preferences and identify effective strategies for reconciling these differences. Pending a final analysis of our adherence data, our findings suggest that providers who utilize decision aids to facilitate SDM for CRC screening should be willing to comply with patient preferences when selecting an appropriate screening test in instances where the patient expresses a desire to participate in the decision-making process, the choice is informed, and there are no mitigating factors to warrant an alternate recommendation. Failure to do so undermines not only the spirit of patient-centered decision making but also the potential effectiveness of SDM as a strategy for increasing uptake of CRC screening.

ACKNOWLEDGMENT

We thank Linda Rosen, MSEE, from the Office of Clinical Research, Boston University School of Medicine, and Daniel Newman, MD, from the Department of Medicine, Boston Medical Center, for their technical support and Jingshun Yang, MS, from the Data Coordinating Center, Boston University School of Public Health, for data management services. Funding support: Agency for Healthcare Research and Quality grant R01HS013912 (PCS).

APPENDIX 1
Knowledge Items*

1. CRC is the number 1 cause of cancer death among nonsmokers. [True]
2. Both men and women are at risk of getting CRC. [True]
3. People 50 years of age and older are more likely to get CRC than younger people. [True]
4. People of all racial and ethnic groups can get CRC. [True]
5. Most colorectal cancers develop from growths called polyps. [True]
6. Removing polyps can prevent CRC. [True]
7. You only have to worry about getting CRC if some one in your family has had it. [False]
8. You can have CRC and not have any symptoms. [True]
9. You can have colorectal polyps and not have any symptoms. [True]
10. The goals of screening are to find polyps and cancer before they cause symptoms. [True]
11. If found early, most colorectal cancers can be cured by surgery. [True]
12. You should begin screening for CRC at age 50. [True]

*Correct responses in brackets.

APPENDIX 2
Satisfaction with the Decision-Making Process Scale[36]

Response frame for questions 1–5:

1. Strongly disagree
2. Disagree
3. Not sure
4. Agree
5. Strongly agree
6. How would you rate the explanations of the screening tests for CRC?
7. How would you rate your doctor's interest in you and your choice of colorectal screening test?
8. How would you rate the reassurance and support offered to you by your doctor?
9. How would you rate the amount of time you had with your doctor?
10. How would you rate the amount of help you got with choosing a CRC screening test?

11. How would you rate the amount of information you got about CRC screening?
12. How would you rate the attention given to what you had to say about CRC screening?

Response frame for questions 6–12:

1 Excellent
2 Very good
3 Good
4 Fair
5 Poor

1. I got as much information as I wanted about colorectal cancer screening.
2. I am satisfied that I was adequately informed about the different tests available for colorectal cancer screening.
3. I had as much input as I wanted in choosing a test for colorectal cancer screening.
4. I am satisfied that my own opinion was important in deciding which colorectal screening test to order.
5. Looking back, I think I relied too much on the opinion of my doctors in deciding which colorectal screening test to order.

REFERENCES

1. Pignone M, Saha S, Hoerger T, Mandelblatt J. Cost-effectiveness analyses of colorectal cancer screening: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med. 2002;137:96–104.

2. Levin B, Lieberman DA, McFarland B, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. CA Cancer J Clin. 2008;58:130–60.

3. US Preventive Services Task Force. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2008;149:627–37.

4. Shapiro JA, Seeff LC, Thompson TD, Nadel MR, Klabunde CN, Vernon SW. Colorectal cancer test use from the 2005 National Health Interview Survey. Cancer Epidemiol Biomarkers Prev. 2008;17:1623–30.

5. Leard LE, Savides TJ, Ganiats TG. Patient preferences for colorectal cancer screening. J Fam Pract. 1997;45:211–8.

6. Wolf AM, Schorling JB. Does informed consent alter elderly patients' preferences for colorectal cancer screening? Results of a randomized trial. J Gen Intern Med. 2000;15:24–30.

7. Pignone M, Bucholtz D, Harris R. Patient preferences for colon cancer screening. J Gen Intern Med. 1999;14:432–7.

8. Pignone M, Harris R, Kinsinger L. Videotape-based decision aid for colon cancer screening: a randomized, controlled trial. Ann Intern Med. 2000;133:761–9.

9. Ling BS, Moskowitz MA, Wachs D, Pearson B, Schroy PC. Attitudes toward colorectal cancer screening tests: a survey of patients and physicians. J Gen Int Med. 2001;16:822–30.

10. Nelson RL, Schwartz A. A survey of individual preference for colorectal cancer screening technique. BMC Cancer. 2004;4:76.

11. Sheikh RA, Kapre S, Calof OM, Ward C, Raina A. Screening preferences for colorectal cancer: a patient demographic study. South Med J. 2004;97:224–30.

12. Dolan JG. Patient priorities in colorectal cancer screening decisions. Health Expect. 2005;8:334–44.

13. Kim J, Whitney A, Hayter S, et al. Development and initial testing of a computer-based patient decision aid to promote colorectal cancer screening for primary care practice. BMC Med Inform Decis Mak. 2005;5:36.

14. Wolf RL, Basch CE, Brouse CH, Shmukler C, Shea S. Patient preferences and adherence to colorectal cancer screening in an urban population. Am J Public Health. 2006;96:809–11.

15. Ruffin MT 4th, Fetters MD, Jimbo M. Preference-based electronic decision aid to promote colorectal cancer screening: results of a randomized controlled trial. Prev Med. 2007;45:267–73.

16. Schroy PC 3rd, Lal S, Glick JT, Robinson PA, Zamor P, Heeren TC. Patient preferences for colorectal cancer screening: how does stool DNA testing fare? Am J Manag Care. 2007;13:393–400.

17. Marshall DA, Johnson FR, Phillips KA, Marshall JK, Thabane L, Kulin NA. Measuring patient preferences for colorectal cancer screening using a choice-format survey. Value Health. 2007;10:415–30.

18. Hawley ST, Volk RJ, Krishnamurthy P, Jibaja-Weiss M, Vernon SW, Kneuper S. Preferences for colorectal cancer screening among racially/ethnically diverse primary care patients. Med Care. 2008;46:S10–6.

19. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med. 1997;44:681–92.

20. Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. Soc Sci Med. 1999;49:651–61.

21. Gravel K, Legare F, Graham ID. Barriers and facilitators to implementing shared decision-making in clinical practice: a systematic review of health professionals' perceptions. Implement Sci. 2006;1:16.

22. Briss P, Rimer B, Reilley B, et al. Promoting informed decisions about cancer screening in communities and healthcare systems. Am J Prev Med. 2004;26:67–80.

23. Braddock CH 3rd, Fihn SD, Levinson W, Jonsen AR, Pearlman RA. How doctors and patients discuss routine clinical decisions: informed decision making in the outpatient setting. J Gen Intern Med. 1997;12:339–45.

24. Braddock CH 3rd, Edwards KA, Hasenberg NM, Laidley TL, Levinson W. Informed decision making in outpatient practice: time to get back to basics. JAMA. 1999;282:2313–20.

25. O'Connor AM, Stacey D, Rovner D, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2001;3:CD001431.

26. Dolan JG, Frisina S. Randomized controlled trial of a patient decision aid for colorectal cancer screening. Med Decis Making. 2002;22:125–39.

27. O'Connor AM, Tugwell P, Wells GA, et al. A decision aid for women considering hormone therapy after menopause: decision support framework and evaluation. Patient Educ Couns. 1998;33: 267–79.

28. O'Connor AM, Drake ER, Fiset V, Graham ID, Laupacis A, Tugwell P. The Ottawa patient decision aids. Eff Clin Pract. 1999; 2:163–70.

29. Winawer SJ, Fletcher RH, Miller L, et al. Colorectal cancer screening: clinical guidelines and rationale. Gastroenterology. 1997;112:594–642.

30. Winawer S, Fletcher R, Rex D, et al. Colorectal cancer screening and surveillance: clinical guidelines and rationale. Update based on new evidence. Gastroenterology. 2003;124: 544–60.

31. Edwards A, Unigwe S, Elwyn G, Hood K. Effects of communicating individual risks in screening programmes: Cochrane systematic review. BMJ. 2003;327:703–9.

32. Nielsen J. Usability Engineering. San Francisco: Morgan Kaufmann Publishers; 1994.

33. Nielsen J. Why you only need to test with 5 users. Available from: URL: http://www.useit.com/alertbox/20000319.html.

34. Strull WM, Lo B, Charles G. Do patients want to participate in medical decision making? JAMA. 1984;252:2990–4.

35. Schroy PC, Glick JT, Robinson P, et al. A cost-effectiveness analysis of subject recruitment strategies in the HIPAA era: results from a colorectal cancer screening adherence trial. Clin Trials. 2009;6:597–609.

36. Levin B, Smith RA, Feldman GE, et al. Promoting early detection tests for colorectal carcinoma and adenomatous polyps: a framework for action. The strategic plan of the National Colorectal Cancer Roundtable. Cancer. 2002;95:1618–28.

37. Massachusetts Department of Public Health. Colorectal cancer. Available from: URL: http://www.mass.gov/dph/cancer/colorectal_cancer.htm.

38. Barry MJ, Cherkin DC, Chang Y, Fowler FJ, Skates S. A randomized trial of a multimedia shared decision-making program for men facing a treatment decision for benign prostatic hyperplasia. Dis Manag Clin Outcomes. 1997;1:5–14.

39. Sutton S, Wardle J, Taylor T, et al. Predictors of attendance in the United Kingdom flexible sigmoidoscopy screening trial. J Med Screen. 2000;7:99–104.

40. McQueen A, Vernon SW, Myers RE, Watts BG, Lee ES, Tilley BC. Correlates and predictors of colorectal cancer screening among male automotive workers. Cancer Epidemiol Biomarkers Prev. 2007;16:500–9.

41. Lieberman DA, Weiss DG, Bond JH, Ahnen DJ, Garewal H, Chejfec G. Use of colonoscopy to screen asymptomatic adults for colorectal cancer: Veterans Affairs Cooperative Study Group 380. N Engl J Med. 2000;343:162–8.

42. Imperiale TF, Wagner DR, Lin CY, Larkin GN, Rogge JD, Ransohoff DF. Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. N Engl J Med. 2000;343:169–74.

43. Lieberman DA, Weiss DG. One-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon. N Engl J Med. 2001;345:555–60.

44. Schoenfeld P, Cash B, Flood A, et al. Colonoscopic screening of average-risk women for colorectal neoplasia. N Engl J Med. 2005;352:2061–8.

45. Cram P, Fendrick AM, Inadomi J, Cowen ME, Carpenter D, Vijan S. The impact of a celebrity promotional campaign on the use of colon cancer screening: the Katie Couric effect. Arch Intern Med. 2003;163:1601–5.

46. Weinstein ND, Atwood K, Puleo E, Fletcher R, Colditz G, Emmons KM. Colon cancer: risk perceptions and risk communication. J Health Commun. 2004;9:53–65.

47. Lafata JE, Divine G, Moon C, Williams LK. Patient-physician colorectal cancer screening discussions and screening use. Am J Prev Med. 2006;31:202–9.

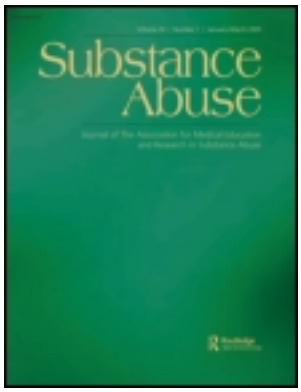
48. Wackerbarth SB, Tarasenko YN, Joyce JM, Haist SA. Physician colorectal cancer screening recommendations: an examination based on informed decision making. Patient Educ Couns. 2007;66:43–50.

49. Ling BS, Trauth JM, Fine MJ, et al. Informed decision-making and colorectal cancer screening: is it occurring in primary care? Med Care. 2008;46:S23–9.

50. Pignone M. Patient preferences for colon cancer screening: the role of out-of-pocket costs. Am J Manag Care. 2007;13:390–1.

51. Griffith JM, Lewis CL, Brenner AR, Pignone MP. The effect of offering different numbers of colorectal cancer screening test options in a decision aid: a pilot randomized trial. BMC Med Inform Decis Mak. 2008;8:4.

52. McQueen A, Bartholomew LK, Greisinger AJ, et al. Behind closed doors: physician-patient discussions about colorectal cancer screening. J Gen Intern Med. 2009;24:1228–35.



Substance Abuse

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wsub20

Alcohol Brand Preferences of Underage Youth: Results from a Pilot Survey Among a National Sample

Michael Siegel MD, MPH [a], William DeJong PhD [a], Timothy S. Naimi MD, MPH [b], Timothy Heeren PhD [c], David L. Rosenbloom PhD [d], Craig Ross MBA [e], Joshua Ostroff BA [e] & David H. Jernigan PhD [f]

[a] Department of Community Health Sciences, Boston University School of Public Health, Boston, Massachusetts, USA

[b] Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

[c] Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA

[d] Join Together, Boston University School of Public Health, Boston, Massachusetts, USA

[e] Virtual Media Resources, Natick, Massachusetts, USA

[f] Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

Available online: 11 Jul 2011

To cite this article: Michael Siegel MD, MPH, William DeJong PhD, Timothy S. Naimi MD, MPH, Timothy Heeren PhD, David L. Rosenbloom PhD, Craig Ross MBA, Joshua Ostroff BA & David H. Jernigan PhD (2011): Alcohol Brand Preferences of Underage Youth: Results from a Pilot Survey Among a National Sample, Substance Abuse, 32:4, 191-201

To link to this article: http://dx.doi.org/10.1080/08897077.2011.601250

Substance Abuse, 32:191–201, 2011

ISSN: 0889-7077 print / 1547-0164 online
DOI: 10.1080/08897077.2011.601250

Routledge
Taylor & Francis Group

Alcohol Brand Preferences of Underage Youth: Results from a Pilot Survey Among a National Sample

Michael Siegel, MD, MPH
William DeJong, PhD
Timothy S. Naimi, MD, MPH
Timothy Heeren, PhD
David L. Rosenbloom, PhD
Craig Ross, MBA
Joshua Ostroff, BA
David H. Jernigan, PhD

ABSTRACT. This study is the first investigation to explore the alcohol brand preferences of underage youth via a national survey. The authors conducted a pilot study of a new, Internet-based alcohol brand survey with 108 youth aged 16 to 20 years who were recruited from an existing panel and had consumed alcohol in the past month. The authors ascertained respondents' consumption of each of 380 alcohol brands during the past 30 days, including which brands of alcohol were consumed during heavy drinking episodes. The findings suggest that, despite the wide variety of alcohol brands consumed by older adolescents in this study, the volume of alcohol consumed is concentrated among a relatively small number of brands. Accurate measurements of alcohol brand preferences will enable important new research into the factors that influence youth drinking behavior. This study establishes the feasibility and validity of a new methodology to determine patterns of brand-specific alcohol consumption among underage drinkers.

KEYWORDS. Adolescents, alcohol, alcohol brands, alcohol use, youth

Michael Siegel and William DeJong are affiliated with the Department of Community Health Sciences, Boston University School of Public Health, Boston, Massachusetts, USA.

Timothy S. Naimi is affiliated with the Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA.

Timothy Heeren is affiliated with the Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA.

David L. Rosenbloom is affiliated with Join Together, Boston University School of Public Health, Boston, Massachusetts, USA.

Craig Ross and Joshua Ostroff are affiliated with Virtual Media Resources, Natick, Massachusetts, USA.

David H. Jernigan is affiliated with the Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.

Address correspondence to: Michael Siegel, MD, MPH, Professor, Department of Community Health Sciences, Boston University School of Public Health, 801 Massachusetts Avenue, 3rd Floor, Boston, MA 02118, USA (E-mail: mbsiegel@bu.edu).

This study was supported by the National Institute on Alcohol Abuse and Alcoholism Center grant P60AA1375905S1.

INTRODUCTION

Alcohol use among adolescents is a major public health problem (1–4). Excessive alcohol consumption contributes to approximately 4600 deaths and 275,000 years of potential life lost among underage youth annually in the United States (5). Adequate surveillance of youth alcohol use is essential to identify the causes of youth drinking in order to plan interventions to prevent its consequences. Although there have been several investigations that examined the types of alcoholic beverages consumed by adolescents (e.g., beer, liquor, wine) (6, 7), we are unaware of any previous studies that reported on the alcohol brands (e.g., Bud Light, Bacardi Silver) consumed by underage youth. This is a critical gap in the literature. Alcohol is marketed and consumed at the brand level, and without any idea of the brands that youth are drinking, it is impossible to develop a complete understanding of the factors—such as alcohol marketing—that influence their drinking behavior.

In 2003, the Institute of Medicine noted this serious flaw in the existing alcohol research literature and recommended the collection of alcohol brand preference data from underage drinkers (4). To date, however, there are no published national data on youth alcohol consumption at the brand level. This paper presents our efforts to fill this glaring deficiency by conducting a pilot study, which we believe is the first survey to comprehensively ascertain youth alcohol brand preferences using a national sample of drinking youth.

Elucidating the brand-specific patterns of alcohol consumption among underage youth would make 4 important contributions. First, identifying the brands popular among young drinkers would allow researchers to examine the relationship between brand-specific advertising exposure and brand-specific alcohol consumption, thus providing the strongest evidence to date regarding whether advertising influences youth drinking. Although previous studies have documented that underage youth are heavily and disproportionately exposed to alcohol advertising for a number of alcohol brands (8–10), without knowing whether youth are actually drinking these brands, we cannot determine whether the advertising is actually affecting their alcohol consumption.

Second, identifying the brands of alcohol that youth consume would greatly enhance our understanding of the factors that influence youth alcohol use. For example, identifying differences in alcohol brands consumed by different age groups and by youth with differing frequencies or intensity of alcohol use may provide insights into the factors that influence the progression of alcohol use behavior.

Third, ascertaining youth alcohol use by brand may result in a more accurate description of drinking behavior among youth. Previous research has established that greater specificity in asking about alcoholic beverage types results in higher self-reported consumption (11, 12). By extension, inquiring about specific alcohol brands could result in the most accurate assessment of youth alcohol consumption to date. In fact, Casswell et al. (13) found that asking subjects to report the brands of alcohol they consume was one of the key factors in their ability to account for 94% of per capita alcohol consumption (as measured by sales data), as compared with less than 60% in prior surveys (14, 15).

Fourth, identifying the patterns of alcohol brand consumption among youth will help establish the feasibility of including alcohol brand use questions on federal or national surveys. A number of national surveys assess cigarette brand preference, which is feasible because cigarette brand use is concentrated among a relatively small number of brands. If alcohol brand consumption is also concentrated among a relatively small number of brands, then assessments of youth alcohol brand preference in national or federal surveys would be highly feasible because only a limited number of brands would need to be listed.

In this paper, we report the results of a pilot study, which we believe is the first national survey dedicated to measuring alcohol brand preferences among underage youth. Although this pilot study has a relatively small sample size, these findings do provide the first national data on youth brand preferences. No previously published paper reports either brand-specific or type-specific alcohol consumption with the comprehensiveness of our survey.

METHODS

Background

A major reason for the absence of studies on alcohol brand use among youth is the lack of an established methodology to collect such data. There are hundreds of major alcohol brands so researchers have assumed that it would take too long to collect such data, making brand research costly and impracticable. Yet, by using a combination of carefully crafted skip patterns, piping questions (using responses from previous questions on brand use to elicit more detailed information on alcohol consumption patterns for the identified brands), and Internet forms that include lists of brands with check boxes, we developed a survey instrument that assesses alcohol brand preferences within a reasonable time frame.

This pilot study was designed to determine our ability to use a preexisting Internet panel to administer the survey to a national sample of underage youth and obtain a valid assessment of their alcohol consumption.

Design

To conduct our survey, we utilized a prerecruited Internet panel developed by Knowledge Networks, the only US company that maintains an Internet panel (the Knowledge Panel) that was created using a national probability sample. The company recruited households to its Knowledge Panel sample through a combination of random digit dialing (RDD) and address-based sampling (ABS), which involves probability sampling of addresses from the US Postal Service's Delivery Sequence File (16).

The Knowledge Networks Internet youth panel provides high survey completion rates because of the ongoing relationship between the youth and the panel staff. To ensure adequate representation of panelists across race/ethnicity, telephone numbers from phone banks with higher concentrations of blacks and Hispanics are oversampled. To ensure adequate participation across levels of socioeconomic status, subjects agreeing to participate in the panel who do not have Internet access are given WebTV and Internet access and training for free.

Previous research has validated the alcohol data derived from adults in the Knowledge Networks Internet panel. Heeren et al. (17) compared the results of an alcohol survey conducted through Knowledge Networks with results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). Estimates of current drinking were similar to those from NESARC, demonstrating that the Knowledge Networks panel is a less expensive, viable alternative to telephone and in-person surveys for assessing drinking behavior.

Sample

Knowledge Networks recruited 108 youth aged 16 to 20 from its existing Internet panel to participate in the study by sending an e-mail invitation. The invitation did not disclose that the survey was related to alcohol consumption. Those who agreed to participate were provided a secure link to access the study site.

The initial screening question asked respondents to report on how many days out of the past 30 they had consumed at least 1 drink of alcohol. A drink was defined as a 12-ounce can or bottle of beer; a 5-ounce glass of wine or champagne; an 8.5-ounce flavored malt beverage; an 8-ounce alcohol energy drink; a 12-ounce wine cooler; 8.5 ounces of malt liquor; 1.5 ounces of liquor (spirits or hard alcohol), whether in a mixed drink or as a shot; and 2.5 ounces of cordials or liqueurs, whether in a mixed drink, a coffee drink, or consumed on their own.

Respondents who had consumed at least 1 drink of alcohol in the past 30 days were provided with an online consent form, which described the details of the study, risks and benefits, and the procedures in place to protect the confidentiality of their responses. Participants who provided informed consent completed the Internet-based questionnaire, which ascertained the alcoholic beverage brands they consumed within the past 30 days. After completion of the survey, a $25 gift was credited to the panel member's account. The protocol was approved by the Institutional Review Board of the Boston University Medical Center. A Certificate of

Confidentiality was obtained from the National Institutes of Health to help protect the confidentiality of the panelists' responses.

Because this was a pilot study, with funding for surveys of only about 100 subjects, we used a consecutive sampling process, enrolling the first 100 adolescents who responded to the e-mail invitation and were found to be eligible after they completed the screening questionnaire. A total of 1028 e-mail invitations were sent out. It took just 1 week to recruit the desired sample. During that week, 360 respondents (35%) completed the screening questionnaire: 108 were qualified (i.e., had consumed at least 1 drink of alcohol in the past 30 days) and completed the survey. We exceeded the desired sample size of 100 because closing the survey does not throw out subjects who are already online.

Survey Instrument

The Internet-based survey instrument was developed to assess brand-specific alcohol consumption among underage youth. The list of assessed alcohol brands was generated using 2 main sources. The first source was the complete list of alcohol brands measured by GfK Mediamark Research & Intelligence (GfK MRI) in its Survey of the Adult Consumer, a written survey of a representative sample of approximately 10,000 US adults. GfK MRI's survey, which ascertains the prevalence of use of various consumer products, inquires about the past-6-month and past-7-day consumption of 90 beer brands and 81 wine brands and the past-6-month and past-30-day consumption of 17 flavored alcoholic beverage brands and 132 spirits brands.

The second source was a list generated by TNS Media Intelligence, which is an advertising industry standard source that monitors advertising occurrences and expenditures in more than 300 national periodicals. The survey instrument included every alcohol brand advertised in the magazines during any of the years 2001 to 2006, except for wine and champagne. Due to the extensive number of advertised wine and champagne brands, it was sufficient and more practical to include only brands advertised in 2006.

Alcoholic energy drinks are not included in the GfK MRI list. To identify these brands, we performed Internet searches and used an extensive list compiled by the National Association of Attorneys General as part of an ongoing investigation into the marketing of these beverages. Finally, we included all alcohol brands reported by participants in the preliminary pilot study that were not on our initial list.

The final survey instrument included 61 brands of beer, 81 brands of wine or champagne, 19 brands of flavored alcoholic beverages (including flavored malt beverages, alcopops, wine coolers, and malt liquor), 35 types of mixed drinks, 38 brands of alcoholic energy drinks, 14 brands of bourbon, 3 brands of brandy, 8 brands of cognac, 9 brands of gin, 19 brands of rum, 15 brands of scotch, 13 brands of tequila, 30 brands of vodka, 8 brands of whiskey, and 27 brands of cordials or liqueurs. In total, the instrument assessed 380 brands of alcoholic beverages. Brand extensions were ascertained as a single brand category if they are typically advertised together (e.g., Absolut flavored vodkas); they were ascertained separately if they are typically advertised separately (e.g., Budweiser, Bud Light).

For each category of alcohol, the respondents checked off which specific brands they had consumed during the past 30 days. If a specific brand was not listed, then respondents entered the name, giving as specific a name as possible. After identifying the brands they had consumed in the past 30 days, the respondents reported the number of days during the past 30 that they had consumed each brand and how many drinks of each brand they usually had on a day when they drank that brand.

Respondents also reported on how many days out of the past 30 they had consumed 5 or more drinks in a row (that is, within a couple of hours). Next, for each category of alcohol, they selected all of the brands, either alone or in combination, which they remembered drinking in the past 30 days on those occasions when they drank at that level.

Analysis

The 3 main data analysis questions were as follows: (1) Of underage youth (aged 16 to 20)

who consumed alcohol in the past 30 days, what proportion reported having consumed each type and each brand of alcohol? (2) Of underage youth who consumed 5 or more drinks in row, what proportion used each type and brand of alcohol during these drinking episodes? (3) What is the overall market share, by volume of alcohol consumed in the past month, for each alcohol brand?

The market share for each alcohol brand by volume consumed was calculated by dividing the total number of drinks consumed during the past 30 days for each brand by the total number of drinks of all brands consumed during the past 30 days. The number of drinks of a brand consumed during the past 30 days was estimated by multiplying the number of days that brand was consumed by the usual number of drinks for that brand on days when it was consumed.

Knowledge Networks applied statistical weighting adjustments to account for selection deviations and to render the sample representative of the underlying population. These weights accounted for the different selection probabilities associated with the RDD- and ABS-based samples, the oversampling of minority communities, nonresponse to panel recruitment, and panel attrition. Poststratification adjustments were based on demographic distributions from the Current Population Survey (CPS) conducted by the US Bureau of the Census.

Validation of Survey Findings

Because there has been no previously published survey of youth alcohol brand preferences, there is no true gold standard to which we can compare our findings. Instead, we can validate our findings against GfK MRI's Survey of the Adult Consumer, a written survey of a representative sample of approximately 10,000 US adults used to ascertain the prevalence of use of various consumer products, including past-30-day consumption of 132 spirits brands. Since GfK MRI defines adults as aged 18 years or older, their findings on types of alcohol used can be applied as a validity check on our own findings for the 18- to 20-year-olds in our Internet panel. We previously obtained data on past-30-day use of various types of liquor among 18- to 20-year-olds from the 2007 Survey of the American Consumer (18), and we used these data here to validate our survey findings.

RESULTS

Description of Sample

The sample was slightly overrepresentative of males (55.6%) and somewhat more representative of older adolescents (age 16: 13.9%; age 17: 19.4%; age 18: 15.7%; age 19: 29.6%; age 20: 21.3%). By race/ethnicity, 66.7% of respondents were non-Hispanic white, 22.2% were Hispanic, 8.3% were black, and 2.8% other race/ethnicity. The mean number of days in the past month on which alcohol was consumed was 4.9, and the median was 3. Approximately two thirds of the respondents (65.7%) reported drinking no more than an average of once per week. The proportion of respondents who consumed 5 or more drinks in a row during the past 30 days was 61.1%.

Feasibility of the Survey Methodology

It was feasible to administer the alcohol brand preference survey using Knowledge Networks' prerecruited Internet panel. We received 108 responses within 1 week, with only 1 e-mail solicitation and no follow-up messages to panelists who did not immediately respond. The median time it took for respondents to complete the survey was 16 minutes; 68.5% of respondents completed the survey in 20 minutes or less. Of 70 respondents who answered an open-ended question regarding whether they had any problems filling out the survey, 65 (92.8%) reported no problems at all. Only 1 respondent reported difficulty with the questions about alcohol brand preferences. That respondent commented that it was difficult to remember specific information about alcohol consumption that occurred 30 days ago. The other 4 respondents reported difficulty with questions related to drinking among relatives, which were not relevant to the present study.

Validation: Comparison of Pilot Survey with GfK MRI Data, 18 to 20 Years of Age

Estimates of beverage category preferences among the 18- to 20-year-old respondents in our study were similar to those from the GfK MRI national survey for most of the alcoholic beverage types (Table 1). The correlation between our estimates and those from GfK MRI was high ($r = .86, P = .0006$), and the correlation between prevalence rank was similarly high (Spearman's rho = .82, $P = .0021$). Overall, the mean absolute difference in prevalence of consumption by beverage category was 1.5%, whereas the mean relative difference was 29.2%.

Alcoholic Beverage Type Preferences of Underage Drinkers, 16 to 20 Years of Age

For our survey respondents, all underage drinkers aged 16 to 20 years, the 3 most popular alcoholic beverage types (based on prevalence of any consumption) were beer (consumed by 67.1%), flavored alcoholic beverages (51.9%), and vodka (43.9%) (Table 2). These were also the types most commonly consumed during episodes when respondents had 5 or more drinks in a row. As a proportion of total drinks consumed, beer accounted for 43.4% of the total alcohol consumed by these youth, followed by flavored alcoholic beverages at 16.5% and rum at 13.0%. The top 5 alcoholic beverage types (beer, flavored alcoholic beverages, rum, vodka, and whiskey) accounted for 90.4% of the total volume of alcohol consumed. All liquor types combined accounted for 55.4% of the total volume of alcohol consumed.

Alcohol Brand Preferences of Underage Drinkers, 16 to 20 Years of Age

Our survey identified 160 brands of alcohol that were consumed in the past 30 days by our sample of underage drinkers. The top 10 brands consisted of 3 beer brands (Bud Light, Budweiser, and Coors Light), 2 brands of flavored alcoholic beverages (Smirnoff Malt Beverages and Mike's Hard Lemonade, Hard Iced Tea, and malt cocktails), 2 brands of rum (Bacardi and Captain Morgan), 1 brand of vodka (Smirnoff), 1 brand of bourbon (Jack Daniels), and 1 brand of tequila (Jose Cuervo) (Table 3). Only 11 brands (the 10 listed previously plus Absolut Vodka) were consumed by 10% or more of the adolescent drinkers in our sample. The brands of alcohol used during episodes when respondents consumed 5 or more drinks in a row were similar to those consumed overall.

TABLE 1. Comparison of Prevalence of Past-30-Day Consumption of Wine and Liquor Types Among 18- to 20-Year-Olds—Knowledge Networks Pilot Study[a] versus GfK MRI Survey of the Adult Consumer[b]

Beverage type	Pilot study, % (rank)	GfK MRI, % (rank)
Vodka	14.6 (1)	10.4 (1)
Rum	8.5 (2)	9.1 (2)
Bourbon	7.1 (3)	5.0 (5)
Tequila	7.0 (4)	7.1 (3)
Whiskey	4.7 (5)	3.4 (9)
Wine	4.5 (6)	4.3 (7)
Cognac	3.4 (7)	4.6 (6)
Cordials or liqueurs	3.2 (8)	7.0 (4)
Gin	2.4 (9)	3.5 (8)
Brandy	1.7 (10)	2.7 (10)
Scotch	1.1 (11)	1.8 (11)

[a]Estimate of overall prevalence of consumption of that alcoholic beverage type in the past 30 days based on weighted responses of 18- to 20-year-old participants in the pilot study. Beer was not included in this validation study because there was no comparable assessment of beer consumption in the past 30 days in the GfK MRI study.

[b]Estimate of overall prevalence of consumption of that alcoholic beverage type in the past 30 days based on weighted responses of 18- to 20-year-old respondents in the 2007 GfK MRI Study of the Adult Consumer (18).

TABLE 2. Alcoholic Beverage Types by Percentage Used in Past 30 Days, Percentage Used When Consuming 5 or More Drinks in a Row, and Overall Market Share by Number of Drinks Consumed in Past 30 Days

	Any use in past 30 days[a] % (95% confidence interval) [*N*]	Use when consuming 5 or more drinks in a row, past 30 days[b] % (95% confidence interval) [*N*]	Market share[c] (%)
Beer	67.1% (54.8–79.4) [73]	69.7% (55.9–83.6) [44]	43.4%
Flavored alcoholic beverages[d]	51.9% (39.0–64.9) [52]	51.2% (35.0–67.5) [29]	16.5%
Vodka	43.9% (30.6–57.2) [40]	45.6% (29.0–62.3) [26]	9.4%
Rum	27.1% (16.8–37.4) [29]	26.6% (13.9–39.4) [19]	13.0%[e]
Tequila	20.2% (8.9–31.6) [19]	10.2% (2.5–17.8) [8]	1.1%
Bourbon	19.8% (7.9–31.7) [18]	25.2% (8.2–42.1) [12]	2.8%
Whiskey	14.6% (3.7–25.4) [9]	15.4% (0.0–30.8) [5]	8.1%[f]
Cordials/liqueurs	13.0% (5.9–20.1) [16]	10.9% (2.0–19.9) [6]	0.6%
Wine	12.3% (1.7–22.8) [10]	6.8% (0.0–13.5) [5]	1.2%
Alcoholic energy drinks	10.9% (4.3–17.4) [14]	5.9% (0.0–12.3) [4]	1.6%
Cognac	8.7% (1.6–15.7) [7]	9.2% (0.8–17.7) [5]	1.8%
Gin	7.5% (1.6–13.5) [7]	5.3% (0.0–11.9) [3]	0.3%
Scotch	5.6% (0.7–10.5) [7]	3.8% (0.0–9.1) [3]	0.1%
Brandy	5.5% (0.1–11.0) [4]	4.1% (0.0–10.0) [2]	0.1%

Note. N = Unweighted number of respondents reporting consumption of that alcohol type during past 30 days (second column) or reporting consumption of that alcohol type when having 5 or more drinks in a row during past 30 days (third column).

[a]Proportion of all respondents who reported drinking that alcoholic beverage type on 1 or more days during the past 30 days.

[b]Proportion of respondents consuming 5 or more drinks in a row who reported drinking that type of alcohol during these episodes. Proportion of respondents who reported drinking 5 or more drinks in a row was 61.1%.

[c]Includes flavored alcoholic beverages, alcopops, malternatives, and malt liquor.

[d]Proportion of total number of drinks of that alcoholic beverage type in the past 30 days among all respondents to the total number of drinks of all alcoholic beverage types in the past 30 days among all respondents.

[e]Market share for rum is higher than for vodka, despite lower prevalence of use, because average monthly number of drinks per respondent is higher (26.9 drinks per month average for rum; 14.1 drinks per month average for vodka).

[f]High market share may be an anomaly due to very high reported consumption of brand during episodes when respondent consumed 5 or more drinks in a row.

We assessed brand market shares by calculating the volume of each brand consumed during the past 30 days. The majority of alcohol consumed was accounted for by a small number of brands, despite the large total number of brands consumed (Table 4). Of the 160 brands consumed, the top 10 brands accounted for 64.7% of the alcohol consumed, the top 20 brands accounted for 80.6%, and the top 30 brands accounted for 88.4%. Only 5 alcohol brands had market shares of 5% or greater and just 21 brands had market shares of 1% or greater.

We found a small concentration of preferred brands within alcoholic beverage categories. For example, although respondents reported drinking 37 different beer brands, the top 5 beer brands (Bud Light, Budweiser, Busch, Coors Light, and Natural Ice) accounted for three fourths (74.9%) of the total volume of beer consumed. The top 10 beer brands accounted for 86.4% of the beer consumed. Similarly, although respondents reported drinking 7 different brands of rum, the top 2 brands (Bacardi and Captain Morgan) accounted for 95.5% of the rum consumed. For tequila, youth reported consuming 7 different brands, but the top 2 brands (Jose Cuervo and Patron) accounted for 87.0% of the tequila consumed. For vodka, there were 18 brands consumed, but the top 5 accounted for 79.6% of overall vodka consumption by volume. Six brands of bourbon were cited by respondents, but 1 brand (Jack Daniels) accounted for 85.5% of consumption by volume. For flavored alcoholic beverages/malt liquor, 19 brands were consumed, but the top 5 brands (Smirnoff Malt Beverages, Mike's, Steel Reserve, Bacardi Silver, and Seagram's Smooth Malt) accounted for 93.5% of consumption by volume.

TABLE 3. Top 10 Leading Alcohol Brands by Percentage Use in Past 30 Days and Percentage Use When Consuming 5 or More Drinks in a Row in Past 30 Days

Any use in past 30 days[a] % (95% confidence interval) [*N*]			Percentage use when consuming 5 or more drinks in a row, past 30 days[b] % (95% confidence interval) [*N*]		
Rank/brand	Type	Percentage use	Rank/brand	Type	Percentage use
1. Bud Light	Beer	33.7% (20.6–46.7) [33]	1. Smirnoff Malt Beverages	Flavored alcoholic beverage	25.6% (12.0–39.2) [15]
2. Smirnoff Malt Beverages	Flavored alcoholic beverage	26.2% (15.6–36.7) [27]	2. Jack Daniels	Bourbon	21.8% (4.8–38.9) [9]
3. Budweiser	Beer	22.6% (10.8–34.4) [18]	3. Bud Light	Beer	21.2% (8.1–34.3) [16]
4. Smirnoff Vodkas	Vodka	18.7% (7.8–29.6) [20]	4. Budweiser	Beer	16.9% (0.9–33.0) [5]
5. Jack Daniels	Bourbon	16.6% (4.9–28.3) [15]	5. Mike's	Flavored alcoholic beverage	14.8% (0.0–29.6) [7]
6. Coors Light	Beer	14.0% (6.9–21.1) [19]	6. Busch	Beer	14.2% (0.0–28.9) [7]
7. Mike's	Flavored alcoholic beverage	13.6% (3.3–23.9) [11]	7. Smirnoff Vodkas	Vodka	14.1% (4.8–23.5) [10]
8. Jose Cuervo	Tequila	13.2% (2.5–23.8) [11]	8. Captain Morgan Rums	Rum	12.8% (3.4–22.2) [9]
9. Bacardi Rums	Rum	12.6% (5.4–19.8) [15]	9. Natural Ice	Beer	12.2% (0.0–27.2) [4]
10. Captain Morgan Rums	Rum	11.7% (4.4–18.9) [13]	10. Bacardi Rums	Rums	11.7% (3.2–20.2) [9]

Note. N = Unweighted number of respondents reporting consumption of that alcohol brand during past 30 days (columns 1 to 3) or reporting consumption of that alcohol brand when having 5 or more drinks in a row during past 30 days (columns 4 to 6).

[a]Proportion of all respondents who reported drinking that alcohol brand on 1 or more days during the past 30 days.

[b]Proportion of respondents consuming 5 or more drinks in a row who reported drinking that brand of alcohol during these episodes. Proportion of respondents who reported consuming 5 or more drinks in a row was 61.1%.

TABLE 4. Top 10 Leading Alcohol Brands by Overall Market Share by Number of Drinks Consumed in Past 30 Days[a]

Rank/brand	Beverage type	Market share, % [*N*]
1. Bud Light	Beer	14.0 [33]
2. Captain Morgan Rums	Rum	10.8 [13]
3. Budweiser	Beer	8.7 [18]
4. Five Star Whiskey	Whiskey	7.8 [1][b]
5. Smirnoff Malt Beverages	Flavored alcoholic beverage	5.2 [27]
6. Mike's	Flavored alcoholic beverage	4.8 [11]
7. Busch	Beer	3.8 [7]
8. Coors Light	Beer	3.6 [19]
9. Steel Reserve	Flavored alcoholic beverage	3.6 [2][b]
10. Jack Daniels	Bourbon	2.4 [15]

Note. N = Unweighted number of respondents reporting consumption of that alcohol brand during past 30 days.
[a]Proportion of total number of drinks of that alcohol brand in the past 30 days among all respondents to the total number of drinks of all brands in the past 30 days among all respondents.
[b]High market share may be an anomaly due to very high reported consumption of brand during episodes when respondent consumed 5 or more drinks in a row.

DISCUSSION

This study has demonstrated the feasibility of comprehensively ascertaining brand-specific alcohol consumption among a national sample of underage youth using a prerecruited Internet panel. The high concordance between our results and estimates of type-specific alcoholic beverage consumption from the GfK MRI Survey of the Adult Consumer confirms the validity of our study methodology in ascertaining type-specific patterns of consumption among underage youth and therefore supports the use of this method for assessing brand-specific alcohol consumption. To the best of our knowledge, this paper is the first to comprehensively report youth alcohol brand consumption patterns among a national sample.

Although the GfK MRI Survey of the Adult Consumer does assess brand-specific alcohol consumption among a small subset of underage drinkers (those aged 18 to 20), it does not include persons below the age of 18 years and thus provides little information about alcohol consumption among the high school and junior high school populations. In addition, the GfK MRI results are not publicly available or publicly reported.

The present research is limited because of the pilot study's small sample size and the relatively low precision of our brand consumption prevalence estimates. Even so, the study results support several important conclusions about brand-specific alcohol consumption among underage youth. First, our findings provide preliminary evidence that, although the spectrum of alcohol brands consumed is wide, the volume of alcohol consumed by underage persons is concentrated in a rather small number of alcohol brands. The top 10 brands alone accounted for 64.7% of all alcoholic beverage drinks consumed by the survey respondents. Only 11 alcohol brands were consumed by 10% or more of respondents. By total volume consumed, fully 88.4% of the alcohol consumption reported by our sample could be accounted for by 30 alcohol brands. If confirmed in a larger sample, this apparent concentration of alcohol brand preferences suggests that it may be feasible for national surveys to ascertain youth alcohol brand preferences quickly and efficiently, something that was not previously thought possible.

Second, there is a small concentration of preferred brands within the alcoholic beverage categories that are widely popular among youth drinkers. For example, although respondents reported drinking 37 different beer brands, the top 5 beer brands (Bud Light, Budweiser, Busch, Coors Light, and Natural Ice) accounted for three fourths (74.9%) of the total volume of beer consumed. Similarly, the top 2 brands of rum accounted for 95.5% of the rum consumed, and

the top 2 brands of tequila accounted for 87.0% of the tequila consumed.

Third, we found that flavored alcoholic beverages are extremely popular among underage drinkers, with more than half reporting consumption of this type of alcoholic beverage during the past month. This is a striking finding, since this type of alcoholic beverage makes up only 2% of the total market (19). This finding demonstrates an important benefit of improved surveillance of the types and brands of alcohol consumed by underage youths: it may point attention to specific brands or beverage categories that are disproportionately popular among youth.

An advantage of this study was that unlike previous surveys, our methodology did not rely upon youth to accurately classify the types of alcoholic beverages that they consume. Our survey specifically assessed consumption of each of the 380 brands. We therefore believe that this method of ascertaining alcohol consumption patterns by type of alcohol may be the most accurate to date, and will enable more accurate studies of the total volume of alcohol consumed by youth drinkers (13).

The major limitation of this study is the small sample size, which resulted in wide confidence intervals around the point estimates for brand-specific alcohol consumption. Therefore, the specific estimates of brand-specific consumption should be interpreted with caution and viewed only as preliminary estimates. The estimates of market share should be interpreted even more cautiously, because they are affected by the imprecision around multiple brands. Despite these limitations, however, the study reveals that alcohol consumption among older adolescents is concentrated around a limited number of highly popular brands. Larger studies that allow these brands to be definitively identified and then examine youth exposure to advertising for these brands should be a priority, as this research would provide strong evidence as to whether alcohol advertising affects youth drinking behavior. The present study therefore supports the need for a large, national survey of youth alcohol brand preferences and suggests that such a study both is feasible and would provide valid alcohol consumption data.

REFERENCES

1. Johnston LD, O'Malley PM, Bachman JG, Schulenberg JE. *Monitoring the Future National Results on Adolescent Drug Use: Overview of Key Findings, 2008*. Bethesda, MD: National Institute on Drug Abuse; 2009. NIH Publication No. 09-7401.

2. Eaton DK, Kann L, Kinchen S, et al. Youth risk behavior surveillance—United States, 2009. *MMWR Surveill Summ.* 2010;59(SS-5):1–142.

3. US Department of Health and Human Services. *The Surgeon General's Call to Action to Prevent and Reduce Underage Drinking*. Rockville, MD: US Department of Health and Human Services, Office of the Surgeon General; 2007.

4. Bonnie RJ, O'Connell ME, eds. Committee on Developing a Strategy to Reduce and Prevent Underage Drinking, National Research Council and Institute of Medicine. *Reducing Underage Drinking: A Collective Responsibility*. Washington, DC: National Academy Press; 2004.

5. Centers for Disease Control and Prevention. *Alcohol-Related Disease Impact (*ARDI*)*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion; 2010. Available at: http://apps.nccd.cdc.gov/ardi/homepage.aspx. Accessed June 23, 2010.

6. Roeber J, Green DL, Meurer KM, et al. Types of alcoholic beverages usually consumed by students in 9th–12th grades—four states, 2005. *MMWR Morb Mortal Wkly Rep.* 2007;56:737–740.

7. Cremeens JL, Miller JW, Nelson DE, Brewer RD. Assessment of source and type of alcohol consumed by high school students: analyses from four states. *J Addict Med.* 2009;3:204–210.

8. Center on Alcohol Marketing and Youth. *Youth Exposure to Alcohol Advertising on Television, 2001 to 2007*. Washington, DC: Center on Alcohol Marketing and Youth, Georgetown University; 2008.

9. Center on Alcohol Marketing and Youth. *It Can Be Done: Reductions in Youth Exposure to Alcohol Advertising in Magazines, 2001–2005*. Washington, DC: Center on Alcohol Marketing and Youth, Georgetown University; 2007.

10. Center on Alcohol Marketing and Youth. *Youth Exposure to Alcohol Advertising on Television and in National Magazines, 2001 to 2006*. Washington, DC: Center on Alcohol Marketing and Youth, Georgetown University; 2007.

11. Serdula MK, Mokdad AH, Byers T, Siegel PZ. Assessing alcohol consumption: beverage-specific versus grouped-beverage questions. *J Stud Alcohol.* 1999;60:99–102.

12. Russell M, Welte JW, Barnes GM. Quantity-frequency measures of alcohol consumption: beverage-specific vs global questions. *Br J Addict.* 1991;86:409–417.

13. Casswell S, Huckle T, Pledger M. Survey data need not underestimate alcohol consumption. *Alcohol Clin Exp Res.* 2002;26:1561–1567.

14. Kerr WC, Greenfield TK. Distribution of alcohol consumption and expenditures and the impact of improved measurement on coverage of alcohol sales in the 2000 National Alcohol Survey. *Alcohol Clin Exp Res.* 2007;31:1714–1722.

15. Midanik L. The validity of self-reported alcohol consumption and alcohol problems: a literature review. *Br J Addict.* 1982;77:357–382.

16. Knowledge Networks. *KnowledgePanel Design Summary*. Menlo Park, CA: Knowledge Networks; 2009.

17. Heeren T, Edwards EM, Dennis JM, Rodkin S, Hingson RW, Rosenbloom DL. A comparison of results from an alcohol survey of a prerecruited Internet panel and the National Epidemiologic Survey on Alcohol and Related Conditions. *Alcohol Clin Exp Res.* 2008;32:222–229.

18. King C, Siegel M, Jernigan DH, et al. Adolescent exposure to alcohol advertising in magazines: an evaluation of advertising placement in relation to underage youth readership. *J Adolesc Health.* 2009;45:626–633.

19. M. Shanken Communications. *The U.S. Beer Market: Impact Databank Review and Forecast. 2009 Edition*. New York: M. Shanken Communications; 2009.

Alcoholic Beverage Preferences and Associated Drinking Patterns and Risk Behaviors Among High School Youth

Michael B. Siegel, MD, MPH, Timothy S. Naimi, MD, MPH, Jennifer L. Cremeens, PhD, MSPH, David E. Nelson, MD, MPH

Background: Very little is known about the types of alcoholic beverages preferred by youth in the U.S. and the relationship between beverage preference and demographic and behavioral characteristics of these youth.

Purpose: To determine the type of alcoholic beverages consumed by adolescent drinkers and how it varies by drinking patterns.

Methods: In 2010, an analysis was performed using 2007 data from the Youth Risk Behavior Survey (YRBS) conducted among public school students in eight states that included a question on the type of alcohol usually consumed. Analysis was restricted to the 7723 youth who reported consuming at least one drink of alcohol in the past 30 days. Beverage type preferences were analyzed by demographic factors, drinking patterns, and other health-risk behaviors. Logistic regression analyses were conducted to examine the correlates of type-specific alcohol consumption.

Results: Liquor was the strongly preferred alcoholic beverage of choice (43.8%), followed by beer (19.2%) and malt beverages (17.4%), with a very low preference for wine (3.7%) or wine coolers (3.4%). A higher preference for liquor or beer was observed among older youth, among those with a riskier pattern of alcohol consumption (e.g., greater frequency of consumption, binge drinking, or drinking and driving), and among youth who engaged in other risk behaviors.

Conclusions: Riskier patterns of drinking and other health-risk behaviors are associated with an increased preference for hard liquor and beer. Improved surveillance of alcoholic beverage preferences among youth will enable a better understanding of the factors related to youth drinking, allowing the development of more effective interventions.

(Am J Prev Med 2011;40(4):419–426)

Introduction

Excessive alcohol consumption contributes to approximately 4600 deaths and 275,000 years of life lost among underage youth annually in the U.S.[1] Despite slight declines in the past decade, almost half of high school–aged youth report past-month alcohol consumption, mostly in the form of binge drinking,[2] and alcohol use among adolescents remains a major public health problem.[2–6] Little is known, however, about the specific types of alcoholic beverages that underage youths consume, how this beverage-specific profile differs by drinking pattern, or what factors predict the type of alcohol that youths consume.

Identifying the types of alcoholic beverages that youth consume would contribute toward a better understanding of the motivating factors underlying underage drinking behavior.[7] There is evidence that preferences for particular types of alcoholic beverages are associated with different drinking patterns.[7–19] Kuntsche et al., for example, have described wine as being consumed in moderation as a social habit, beer and spirits as most often being used to get drunk, and alcopops as occupying a middle ground.[7] Other studies have identified spirits consump-

From the Department of Community Health Sciences, Boston University School of Public Health (Siegel); Section of General Internal Medicine, Boston University School of Medicine (Naimi), Boston, Massachusetts; the Department of Health Education and Promotion, College of Health & Human Performance, East Carolina University (Cremeens), Greenville, North Carolina; and the National Cancer Institute, NIH (Nelson), Rockville, Maryland

Address correspondence to: Michael B. Siegel, MD, MPH, Department of Community Health Sciences, Boston University School of Public Health, 801 Massachusetts Avenue, 3rd Floor, Boston MA 02118. E-mail: mbsiegel@bu.edu.

0749-3797/$17.00
doi: 10.1016/j.amepre.2010.12.011

tion to be related to a desire to feel the effects of alcohol quickly,[8,19] whereas beer consumption has been associated with risky drinking, including binge drinking, heavy episodic drinking, and drunk driving.[9,11–15]

Because numerous evidence-based prevention strategies, including excise tax policy and alcohol sales and marketing regulation, are beverage-specific,[10] understanding the specific types of alcoholic beverages consumed by young people could also inform the development of appropriate beverage-specific policy and practice interventions. In addition, understanding the relationship between the types of alcoholic beverages that youth prefer and the alcohol source, drinking location, context, and relationship with other health-risk behaviors would provide clues as to the factors that influence youth drinking behavior.

Although several studies have investigated adolescent consumption of various types of alcoholic beverages (e.g., wine, beer, spirits) in other countries,[7–9] there is a paucity of such data in the U.S. There have been only two published studies on type-specific alcoholic beverage consumption among adolescents in the U.S.[20,21] In one study,[20] Roeber et al. reported type-specific consumption of alcoholic beverages among 9th–12th-grade students in four states in 2005 based on the Youth Risk Behavior Survey (YRBS). In the second study,[21] Cremeens et al. used the same data set to examine the type of alcohol consumed and its relationship to binge drinking behavior. However, neither study assessed whether beverage choice was related to drinking context or other personal health-risk behaviors and neither reported the relationship between type-specific consumption and the location of drinking or the source of alcohol.

In this paper, data from the 2007 Youth Risk Behavior Survey in eight states were used to examine beverage-specific drinking patterns among U.S. adolescents. The purposes of the present study were to determine (1) the types of alcoholic beverages preferred by adolescents and how these beverage preferences differ among demographic groups and (2) how alcoholic beverage preferences differ by age, drinking patterns, and other health-risk behaviors.

Methods

Overview

Data were analyzed from the 2007 state Youth Risk Behavior Survey (YRBS), a school-based questionnaire survey of 9th–12th-grade students, in eight states: Arkansas, Florida, Georgia, Hawaii, New Mexico, North Dakota, Utah, and Vermont. Each of these states included an additional question in its 2007 survey that ascertained the type of alcoholic beverage usually consumed by respondents who reported drinking alcohol in the past 30 days.

Sampling

In 2007, these states conducted their survey using a two-stage cluster sample design in order to produce data representative of the state's public school students in Grades 9–12.[3] The first level of clustering was the school level. All public schools were included in the sampling frame. Schools were selected such that their probability of selection was proportional to their enrollment in Grades 9–12.[22] The second sampling frame consisted of classes within the chosen schools. All classes in a required subject or all classes meeting during a particular time period were included in the sampling frame. Equal probability sampling of classes was conducted and all students in selected classes were eligible to participate.

The student sample sizes for the eight states ranged from 1191 (Hawaii) to 8453 (Vermont), with a total sample size of 24,622 across the eight states. School response rates ranged from 76% (Arkansas) to 100% (Vermont), and student response rates ranged from 63% (Hawaii) to 89% (Georgia), resulting in overall response rates ranging from 60% to 82%.[3]

Analyses were restricted to respondents who had consumed at least one alcoholic beverage during the past 30 days (n=8694), which represented 37.8% of the total sample. After deleting respondents who failed to answer the question about alcoholic beverage type preference, the total sample size was 7723.

Data were weighted to represent each state's public school student population for Grades 9–12. Individual sample weights were applied to each record to adjust for student nonresponse and poststratification to the gender, race/ethnicity, and grade-level distribution in that state.

Measures

Alcohol beverage preference was assessed among current drinkers, who were defined as respondents who indicated having had at least one drink of alcohol on at least 1 day during the 30 days prior to survey administration. Alcoholic beverage type was assessed with the question *During the past 30 days, what type of alcohol did you usually drink?* The possible responses were *beer*; *liquor (such as vodka, rum, scotch, bourbon, or whiskey)*; *wine*; *wine coolers (such as Bartles and James or Seagrams)*; *malt beverages (such as Smirnoff Ice, Bacardi Silver, or Hard Lemonade)*; *some other type*; or *no usual type*.

Beverage type preferences were analyzed by state; age; grade; gender; race/ethnicity; frequency of binge drinking (defined as having five or more drinks of alcohol in a row during the past 30 days); frequency of drinking; and driving after drinking (defined as having driven a car shortly after drinking). Beverage type preferences also were analyzed in relation to the usual source of alcohol and the location where alcohol is usually consumed.

Alcohol beverage preferences were analyzed in relation to the following other health-risk behaviors: wearing a seat belt, riding in a car driven by someone who had been drinking alcohol, carrying a weapon, being in a physical fight, feeling helpless, considering suicide, current smoking, marijuana use, TV viewing, number of sexual partners, and unprotected sex. TV viewing was included as a potential correlate of alcoholic beverage type preference because it has been shown to be related to youth alcohol consumption, possibly because of alcohol marketing and the positive portrayal of drinking on TV.[23–25]

Analysis

Analyses were conducted in 2010 with SAS, version 9.1, using procedures that account for the complex survey design and that allowed weighting of the data to produce estimates that were representative of the state school student populations.[26] The analysis relied on a Taylor series variance estimation, which accounts for the clustering and stratification in the survey sampling design.[27]

Both state-specific and pooled analyses across all eight states were conducted. For pooled analyses, an additional weight was employed in order to account for the different student sample sizes in each state and for the differences in state populations in the age range of the sample. One weighting factor accounted for the sample size in each state in order to ensure that responses from each state had an equal influence on the estimation of pooled proportions. A second weighting factor accounted for the population of those aged 10–19 years in each state (based on 2007 age-specific population projections from the U.S. Census Bureau[28]) in order to ensure that states with larger populations had the proper proportionate influence on pooled estimates. By using these two weighting factors, pooled estimates were designed to be representative of the combined population of these eight states.

Logistic regression analyses were conducted to examine the correlates of type-specific alcohol consumption. The outcome variable in each regression was preference for a particular type of alcoholic beverage (i.e., liquor, beer, malt beverages, wine coolers, wine, and beer/liquor combined) compared to all other categories (including "other" and "no usual type"). The independent variables included age, gender, race/ethnicity, usual source of alcohol, usual location of drinking, frequency of drinking, and frequency of binge drinking. All independent variables were included in each model as the aim was to investigate the independent effect of each of these variables on alcoholic beverage type preferences. The other health behaviors were not included in these models in order to avoid multicollinearity, which could have invalidated the findings.

Results

In each state except North Dakota, liquor was the most prevalent type of alcohol usually consumed by 9th–12th-grade students in 2007 (median=43.7%, range=33.9%–45.8%; Appendix A, available online at www.ajpm-online.net). Beer was generally the second most prevalent type of alcohol consumed (median=22.7%, range=17.4%–35.9%), followed closely by malt beverages (median=16.4%, range=12.4%–22.4%). Wine and wine coolers were not reported as the usual alcoholic beverage consumed by more than 4.3% of the youth in any state.

For pooled state data, liquor was the strongly preferred alcoholic beverage of choice (43.8%), followed by beer (19.2%) and malt beverages (17.4%; Table 1). Boys were more likely to prefer liquor and beer, whereas girls were more likely to prefer malt beverages, wine coolers, and wine. Older age was associated with increasing preference for liquor and beer and decreasing preference for malt beverages and wine coolers. Black adolescents were much more likely to prefer malt beverages and much less likely to prefer beer compared with those of any other race or ethnicity.

A riskier pattern of alcohol consumption (both for frequency of drinking and of binge drinking) was associated with an increased preference for liquor and beer and a decreased preference for other beverages (Table 1). Driving after drinking was associated with increased beer consumption and decreased consumption of malt beverages, wine coolers, and wine. Preferences for alcoholic beverages type by usual source of alcohol and usual location of drinking tended to mirror the overall preference pattern in the sample.

The use of other drugs (cigarettes and marijuana) was associated with an increased preference for liquor and beer and a decreased preference for malt beverages, wine, and wine coolers (Table 2). In general, youth who engaged in other risk behaviors were more likely to usually consume liquor. For example, 48.4% of the youth who reported having been in a physical fight preferred liquor, compared to 40.0% of the youth who had not been in a physical fight. Nearly half (49.9%) of the youth who carried a weapon preferred liquor compared to 40.8% of the youth who did not carry a weapon.

In regression models, older adolescents were significantly more likely to report usually drinking beer (OR=1.60, 95% CI=1.002, 2.56, for youth aged ≥18 years compared to those aged 12–14 years), and significantly less likely to report usually drinking malt beverages or wine coolers (Appendix B, available online at ajpm-online.net). Girls were less likely to prefer beer and more likely to prefer malt beverages and wine coolers, as were black adolescents. Hispanic youth were more likely than white youth to prefer malt beverages, and less likely to prefer liquor or beer.

Usual source of alcohol was associated significantly with the type of alcohol usually consumed for only a few variables. Compared to youth whose usual source of alcohol was *other*, those whose usual source was buying alcohol in a store were more likely to usually drink beer (OR=1.81; 95% CI=1.14, 2.87). Youths whose usual source of alcohol was a bar or restaurant were less likely to usually drink wine coolers. Youths whose usual source of alcohol was to have someone buy alcohol for them were more likely to usually drink malt beverages.

Compared to youths whose usual location of drinking was *other*, youths whose usual drinking location was in the home were less likely to usually drink liquor (OR=0.46, 95% CI=0.29, 0.74). Those who usually drank alcohol in another person's home or in a public place were less likely to prefer liquor and more likely to prefer beer.

A high frequency of drinking was associated with a preference for liquor and a decreased preference for malt

Table 1. Usual alcoholic beverage type[a] among 9th- to 12th-grade students by demographic and drinking behavior characteristics—eight states combined,[b] % (SE)

Characteristic (%)	Liquor	Beer	Malt beverages	Wine coolers	Wine	Other	No usual type
Overall (*n*=7723)	43.8 (1.1)	19.2 (1.0)	17.4 (0.7)	3.4 (0.5)	3.7 (0.4)	3.5 (0.4)	9.1 (0.6)
Gender							
Male (49.6)	47.1 (1.6)	25.1 (1.6)	11.1 (0.8)	1.5 (0.5)	2.8 (0.6)	2.7 (0.4)	9.6 (1.0)
Female (74.7)	40.5 (1.6)	13.3 (1.0)	23.7 (1.1)	5.2 (0.7)	4.5 (0.6)	4.1 (0.6)	8.7 (0.7)
Age (years)							
12–14 (7.4)	39.0 (3.9)	17.0 (2.8)	20.0 (2.8)	6.8 (2.6)	2.9 (1.1)	5.3 (1.7)	9.0 (2.2)
15–17 (74.7)	43.9 (1.2)	17.4 (1.1)	18.1 (0.9)	3.4 (0.5)	3.8 (0.5)	3.6 (0.4)	9.8 (0.8)
≥18 (18.0)	44.9 (2.4)	27.2 (2.4)	13.5 (1.7)	2.1 (0.9)	3.2 (0.8)	2.3 (0.7)	6.6 (1.5)
Grade							
9 (25.1)	44.0 (1.8)	12.7 (1.3)	19.7 (1.7)	5.9 (1.2)	3.9 (0.8)	4.7 (0.8)	9.0 (1.2)
10 (25.1)	43.3 (2.5)	18.2 (1.8)	18.4 (1.6)	2.9 (0.6)	3.4 (0.7)	4.3 (0.8)	9.5 (1.1)
11 (24.7)	42.4 (1.8)	19.7 (1.9)	16.4 (1.3)	3.1 (0.7)	4.2 (0.8)	2.9 (0.5)	11.3 (1.3)
12 (25.1)	45.5 (2.4)	25.8 (1.9)	15.4 (1.5)	1.9 (0.6)	3.0 (0.6)	2.0 (0.5)	6.4 (1.3)
Race/ethnicity							
White (59.3)	46.4 (1.4)	24.0 (1.3)	13.7 (1.0)	2.0 (0.3)	3.1 (0.5)	1.8 (0.4)	8.9 (0.9)
Black (19.3)	38.0 (2.6)	5.5 (1.1)	27.3 (2.5)	9.1 (1.5)	4.0 (1.0)	6.4 (0.9)	9.7 (1.8)
Hispanic (16.8)	39.7 (2.0)	17.1 (2.0)	18.9 (1.8)	2.0 (0.6)	5.0 (1.0)	6.9 (1.2)	10.4 (1.6)
Other (4.5)	48.4 (4.6)	13.8 (2.5)	17.3 (2.8)	3.7 (1.6)	4.1 (1.7)	6.1 (2.2)	6.7 (1.9)
Frequency of drinking (days)[c]							
1 or 2 (45.9)	37.2 (1.5)	15.1 (1.1)	23.5 (1.0)	5.3 (0.7)	5.2 (0.7)	4.4 (0.6)	9.2 (0.8)
3–9 (39.2)	46.6 (1.6)	22.4 (1.6)	14.2 (1.1)	1.8 (0.4)	2.6 (0.6)	2.6 (0.5)	9.8 (1.2)
≥10 (15.0)	56.4 (2.7)	23.3 (2.4)	6.8 (1.2)	1.7 (0.9)	1.7 (0.6)	3.0 (0.8)	7.0 (1.3)
Frequency of binge drinking (days)[c]							
None (44.1)	33.6 (1.5)	14.6 (1.1)	25.6 (1.1)	6.1 (0.8)	6.6 (0.7)	4.4 (0.6)	9.2 (1.0)
1 or 2 (31.5)	50.0 (1.8)	18.5 (1.4)	14.0 (1.0)	1.9 (0.4)	2.2 (0.6)	3.4 (0.6)	10.0 (1.2)
≥3 (24.5)	54.2 (2.5)	28.0 (2.3)	7.0 (1.1)	0.5 (0.4)	0.4 (0.2)	2.0 (0.5)	7.9 (1.0)
Driving after drinking[c]							
No (75.9)	43.5 (1.2)	16.0 (0.9)	19.2 (0.8)	3.7 (0.5)	4.3 (0.5)	3.8 (0.5)	9.6 (0.8)
Yes (24.1)	44.4 (2.1)	29.5 (2.1)	11.7 (1.4)	2.5 (0.6)	1.7 (0.5)	2.6 (0.7)	7.7 (1.0)
Usual source of alcohol[d]							
Store (7.2)	45.4 (3.7)	31.9 (4.0)	10.4 (2.5)	2.5 (1.6)	2.8 (1.5)	2.8 (1.4)	4.1 (1.4)
Restaurant or bar (2.1)	51.2 (6.8)	11.3 (4.2)	17.2 (4.4)	0.0 (0.0)	3.6 (2.4)	4.1 (2.7)	12.6 (5.1)
Someone bought it for me (21.4)	47.1 (2.3)	23.6 (2.3)	16.8 (1.7)	1.3 (0.5)	1.8 (0.7)	1.8 (0.6)	7.5 (1.3)
Someone gave it to me (40.5)	39.9 (1.5)	18.4 (1.3)	20.6 (1.2)	3.9 (0.7)	3.5 (0.6)	3.4 (0.6)	10.3 (1.2)
Took it (9.5)	46.6 (3.0)	11.2 (2.3)	17.4 (2.8)	7.1 (1.5)	5.0 (1.5)	4.1 (1.1)	8.5 (1.6)
Other (19.4)	45.5 (2.2)	16.3 (1.6)	13.8 (1.8)	3.4 (0.8)	5.7 (1.1)	4.9 (0.8)	10.4 (1.3)

(continued on next page)

Table 1. *(continued)*

Characteristic (%)	Liquor	Beer	Malt beverages	Wine coolers	Wine	Other	No usual type
Usual location of drinking[e]							
My home (26.8)	38.4 (1.3)	13.9 (1.4)	20.2 (1.2)	5.4 (0.8)	8.8 (1.1)	4.2 (0.8)	9.1 (1.1)
Another person's home (55.9)	45.4 (1.5)	21.8 (1.4)	17.4 (1.0)	2.2 (0.4)	1.7 (0.3)	2.2 (0.3)	9.3 (0.9)
Restaurant or bar (5.3)	49.0 (5.1)	9.8 (2.3)	14.7 (3.3)	6.6 (1.8)	4.5 (1.8)	6.9 (1.9)	8.5 (2.1)
Public place (8.0)	37.9 (3.4)	29.1 (4.1)	13.9 (2.0)	2.3 (0.6)	1.1 (0.6)	5.2 (1.6)	10.5 (2.8)
Other (4.1)	62.5 (4.8)	9.8 (2.6)	9.3 (2.5)	4.8 (2.4)	2.1 (1.4)	7.8 (2.4)	3.7 (1.6)

Note: Weighted percentages of respondents in each stratum are shown in parentheses following the stratum label.
[a]The analysis is restricted to students who reported drinking at least one drink of alcohol on ≥1 days during the past 30 days; based on the question *During the past 30 days, what type of alcohol did you usually drink?*
[b]Data are from the Youth Risk Behavior Survey and are weighted to reflect student nonresponse and to provide estimates representative of the student populations of each state. Additional weighting factors were added to reflect the different sample sizes and student populations in each state. Data presented are representative of the overall student populations of all eight states combined.
[c]Data reflect the past 30 days. These variables were recoded to create three relatively similar sized groups, so that the SEs would not be large in any one group and to create categories based on our conceptualization of the meaning of low, medium, and high frequency of drinking and binge drinking. The original categories for frequency of drinking were 1 or 2 days, 3–5 days, 6–9 days, 10–19 days, 20–29 days, and all 30 days. The original categories for frequency of binge drinking were 0 days, 1 day, 2 days, 3–5 days, 6–9 days, 10–19 days, and ≥20 days.
[d]*Other* includes source of alcohol reported as *at a public event* or *other*.
[e]*Other* includes usual location of drinking reported as *at a public event*, *on school property*, or *in a car*; question not asked in Hawaii, North Dakota, or Vermont

beverages. There was a strong relationship between binge drinking and the usual type of alcohol consumed. Binge drinking was associated with a preference for liquor and a decreased preference for malt beverages, wine coolers, and wine.

Discussion

The present research advances the literature by providing the largest sample to date in which adolescent beverage preferences are measured. It is also the first population-based study to examine alcohol beverage preference among youth based on location of consumption, source of alcohol, and as a function of other health-risk behaviors. Liquor was the most popular beverage preference among almost half of youth drinkers, was almost twice as popular as the next most popular beverage category, was the beverage of choice in seven of eight states, and was the predominant beverage choice for virtually all demographic strata. Furthermore, liquor was disproportionately popular among those who were frequent drinkers or binge drinkers, and among those who reported a variety of other health-risk behaviors (e.g., drinking and driving, carrying a weapon, smoking, having multiple sexual partners or not using condoms).

Beer and malt-based flavored beverages (referred to as malt beverages in the current study) were the second and third most popular alcohol beverage choices among youth. Although beer was a slightly more popular beverage than malt beverages for those drinking more frequently, binge drinking, or drinking and driving, it was nonetheless surprising that malt beverages had a similar overall popularity as beer given the large advertising expenditures for beer relative to malt beverages.[29,30] Data from the Monitoring the Future Study demonstrates that the past-30-day prevalence of beer consumption, which was formerly the clear-cut alcoholic beverage of choice among youth, decreased from 47% to 34% between 1990 and 2008.[31] In many respects, beer and malt beverages are similar in that malt beverages are taxed like beer in most states, and are similar to beer in terms of their distribution patterns and sales venues. Viewed in this light, it may be that malt beverage split the portion of the youth alcohol market that formerly consisted of beer. On the other hand, many malt beverages are highly flavored, and therefore likely more appealing to youth tastes, particularly to those who are relatively new to alcohol consumption or who are female. Because the prevalence of alcohol consumption and binge drinking among girls has increased and is now similar to that of boys, this may partly account for gains for malt beverages compared with beer. Finally, malt beverages are often produced by liquor companies and may be used to build brand loyalty among those who may go on to drink liquor at a later stage in their drinking trajectories. Moreover, the marketing activity for liquor has increased dramatically in the past decade, particularly on cable TV and in media venues whose audiences are disproportionately youthful relative to the general population.[32]

An important finding of this paper is the pattern of a decreased preference for malt beverages and wine coolers

Table 2. Usual alcoholic beverage type[a] among 9th- to 12th-grade students by other health-risk behaviors—eight states combined,[b] % (SE)

Characteristic (%)	Liquor	Beer	Malt beverages	Wine coolers	Wine	Other	No usual type
Overall (*n*=7723)	43.8 (1.1)	19.2 (1.0)	17.4 (0.7)	3.4 (0.5)	3.7 (0.4)	3.5 (0.4)	9.1 (0.6)
Wear seat belt[c]							
Sometimes, mostly, always (84.8)	43.0 (1.2)	18.8 (1.0)	18.0 (0.7)	3.7 (0.5)	3.8 (0.4)	3.2 (0.4)	9.6 (0.8)
Never or rarely (15.3)	49.2 (2.3)	21.3 (2.4)	13.4 (2.1)	1.9 (1.0)	2.9 (0.9)	4.9 (1.0)	6.3 (1.2)
Rode in car with drinking driver[d]							
No (53.7)	42.5 (1.5)	17.0 (1.1)	19.2 (0.9)	3.2 (0.5)	5.1 (0.6)	3.5 (0.5)	9.6 (0.9)
Yes (46.3)	45.2 (1.5)	21.8 (1.7)	15.3 (1.1)	3.6 (0.7)	2.0 (0.4)	3.5 (0.5)	8.7 (0.9)
Carried a weapon[e]							
No (71.7)	40.8 (1.3)	19.5 (1.1)	19.8 (0.9)	3.4 (0.5)	4.0 (0.5)	2.9 (0.4)	9.5 (0.7)
Yes (28.4)	49.9 (1.9)	19.7 (1.6)	11.8 (1.2)	3.4 (1.0)	2.5 (0.6)	4.3 (0.6)	8.4 (1.1)
Have been in physical fight[f]							
No (55.2)	40.0 (1.6)	21.2 (1.2)	19.1 (1.2)	3.1 (0.5)	3.6 (0.4)	2.5 (0.4)	10.5 (1.0)
Yes (44.8)	48.4 (1.6)	17.1 (1.6)	15.2 (1.1)	3.7 (0.7)	3.8 (0.8)	4.7 (0.6)	7.2 (0.9)
Have felt sad or hopeless[g]							
No (65.1)	42.7 (1.4)	22.1 (1.2)	16.3 (0.8)	3.1 (0.6)	2.7 (0.4)	3.3 (0.4)	9.8 (0.8)
Yes (34.9)	45.8 (1.7)	13.7 (1.4)	19.3 (1.4)	4.0 (0.7)	5.4 (0.8)	3.8 (0.7)	8.0 (0.9)
Seriously considered suicide[h]							
No (81.2)	43.1 (1.4)	20.7 (1.1)	17.7 (0.1)	3.4 (0.5)	3.1 (0.4)	3.3 (0.4)	8.9 (0.8)
Yes (18.8)	47.3 (2.0)	13.0 (1.7)	15.4 (1.6)	3.5 (0.9)	6.1 (1.2)	4.4 (0.9)	10.5 (1.4)
Current smoker[i]							
No (63.6)	41.0 (1.6)	16.4 (1.2)	20.2 (1.0)	4.3 (0.7)	4.8 (0.5)	3.6 (0.4)	9.7 (0.9)
Yes (36.4)	47.9 (1.8)	24.4 (1.7)	13.1 (1.2)	1.7 (0.5)	1.7 (0.5)	3.0 (0.6)	8.1 (1.0)
Used marijuana[j]							
No (60.7)	40.5 (1.4)	17.0 (1.1)	20.4 (0.5)	3.8 (0.5)	5.1 (0.5)	3.7 (0.5)	9.7 (0.8)
Yes (39.4)	49.1 (1.6)	22.6 (1.6)	12.9 (1.0)	2.5 (0.6)	1.3 (0.4)	3.0 (0.5)	8.6 (0.9)
Hours of TV per day[k]							
<3 (63.0)	43.3 (1.5)	21.7 (1.3)	16.4 (0.8)	2.2 (0.3)	3.9 (0.5)	2.8 (0.4)	9.7 (0.9)
≥3 (37.0)	44.4 (1.9)	14.0 (1.3)	19.7 (1.5)	5.4 (0.9)	3.5 (0.6)	4.7 (0.6)	8.4 (1.0)
Number of sexual partners[l]							
None (46.6)	39.7 (2.2)	19.7 (1.7)	19.0 (1.5)	2.9 (0.6)	5.0 (0.7)	3.5 (0.6)	10.2 (1.1)
1 (35.5)	44.9 (1.9)	22.0 (2.0)	16.2 (1.6)	2.5 (0.7)	3.3 (0.7)	2.4 (0.6)	8.7 (1.4)
≥2 (17.9)	52.3 (3.4)	17.2 (2.7)	14.1 (2.3)	5.3 (1.9)	0.9 (0.6)	5.5 (1.5)	4.7 (1.2)

(continued on next page)

among older adolescents, accompanied by an increased preference for liquor and beer. The same pattern—a preference for beer and liquor instead of malt beverages and wine coolers—was associated with more frequent and riskier drinking behavior and with other types of addictive behaviors (i.e., smoking and marijuana use). This finding has potential implications for prevention, as it suggests that interventions aimed at young adolescents who drink wine coolers and malt beverages might help avert this transition to more hard-core alcohol and the

Table 2. *(continued)*

Characteristic (%)	Liquor	Beer	Malt beverages	Wine coolers	Wine	Other	No usual type
Had sex without condom[m]							
No (77.5)	42.1 (1.5)	19.5 (1.3)	18.3 (1.1)	3.7 (0.7)	3.8 (0.5)	3.4 (0.5)	9.2 (0.8)
Yes (22.5)	50.3 (2.4)	21.6 (2.3)	13.4 (1.8)	1.7 (0.7)	2.8 (0.9)	3.1 (0.9)	7.1 (1.6)

Note: Weighted percentages of respondents in each stratum are shown in parentheses following the stratum label.
[a]The analysis is restricted to students who reported drinking at least one drink of alcohol on ≥1 days during the past 30 days; based on the question *During the past 30 days, what type of alcohol did you usually drink?*
[b]Data are from the Youth Risk Behavior Survey. Data are weighted to reflect student nonresponse and to provide estimates representative of the student populations of each state. Additional weighting factors were added to reflect the different sample sizes and student populations in each state. Data presented are representative of the overall student populations of all eight states combined.
[c]*How often do you wear a seat belt when riding in a car driven by someone else?*
[d]*During the past 30 days, how many times did you ride in a car or other vehicle driven by someone who had been drinking alcohol?*
[e]*During the past 30 days, how many times did you drive a car or other vehicle when you had been drinking alcohol?*
[f]*During the past 12 months, how many times were you in a physical fight?*
[g]*During the past 12 months, did you ever feel so sad or hopeless almost every day for two weeks or more in a row that you stopped doing some usual activities?*
[h]*During the past 12 months, did you ever seriously consider attempting suicide?*
[i]Current smokers were respondents who reported having smoked ≥1 day during the past 30 days.
[j]*During the past 30 days, how many times did you use marijuana?*
[k]*On an average school day, how many hours do you watch TV?*
[l]*During the past 3 months, with how many people did you have sexual intercourse?*; question not asked in Georgia or Utah
[m]*The last time you had sexual intercourse, did you or your partner use a condom?*; question not asked in Georgia or Utah

associated adoption of heavier and riskier drinking patterns.

This study is subject to several limitations. First, underage youth may under-report how much they drink, although there are no data to suggest that this would skew self-reports of which alcoholic beverages actually were consumed. Second, these data are representative of students from public high schools and not necessarily representative of those who do not attend school or those who attend alternative schools or private schools. Furthermore, the study was a population-based study of eight states and was therefore not nationally representative. Third, there may have been some misclassification with respect to which beverage types were consumed.

Improved surveillance for beverage and brand preferences among youth will enable better understanding of how these factors related to youth drinking, particularly when these data are combined with data about price, availability, and access-related factors experienced by youth who drink alcohol. This will be an important next step to designing and promulgating additional interventions to reduce youth drinking and its consequences.

The authors acknowledge Dr. Robert Brewer, head of the Alcohol Team at the CDC, who led efforts to improve the surveillance of type-specific alcohol use and helped conceptualize the analytic design for the paper. We also acknowledge the Youth Risk Behavior Survey coordinators in the eight states involved in this study: Dave Randall (Hawaii); Kathleen Courtney (Arkansas); Erika Edwards (Vermont); Dan Green, Jim Roeber, and Dr. Kristine Meurer (New Mexico); Jamie Weitz (Florida); Dr. Daphna Kanny (Georgia); Andrea Pena (North Dakota); and Frank Wojtech (Utah).

No financial disclosures were reported by the authors of this paper.

References

1. CDC. Alcohol-related disease impact (ARDI). Atlanta GA: CDC, National Center for Chronic Disease Prevention and Health Promotion. apps.nccd.cdc.gov/ardi/homepage.aspx.
2. Johnston LD, O'Malley PM, Bachman JG, Schulenberg JE. Monitoring the Future national results on adolescent drug use: overview of key findings, 2008. Bethesda MD: National Institute on Drug Abuse, 2009. NIH Publication No. 09-7401.
3. Eaton DK, Kann L, Kinchen S, et al. Centers for Disease Control and Prevention Youth risk behavior surveillance, U.S., 2007. MMWR Surveill Summ 2008;57(4):1–131.
4. Newes-Adeyi G, Chen CM, Williams GD, Faden VB. Surveillance Report #74: trends in underage drinking in the U.S., 1991–2003. Bethesda MD: USDHHS, Public Health Service, NIH, National Institute on Alcohol Abuse and Alcoholism, 2005.
5. USDHHS. The Surgeon General's call to action to prevent and reduce underage drinking. Rockville MD: USDHHS, Office of the Surgeon General, 2007.
6. Bonnie RJ, O'Connell ME, eds. Committee on Developing a Strategy to Reduce and Prevent Underage Drinking, National Research Council and IOM. Reducing underage drinking: a collective responsibility. Washington DC: National Academy Press, 2004.
7. Kuntsche E, Knibbe R, Gmel G, Engels R. "I drink spirits to get drunk and block out my problems. . .": beverage preference, drinking motives and alcohol use in adolescence. Alcohol Alcohol 2006;41(5):566–73.
8. Smart RG, Walsh GW. Do some types of alcoholic beverages lead to more problems for adolescents? J Stud Alcohol 1995;56(1):35–8.

9. Kuntsche EN. Radicalisation? An interpretation model of adolescent alcohol use from 1986 to 1988 in Switzerland. Sucht 2001;47(6):393–403.
10. Naimi TS, Brewer RD, Miller JW, Okoro C, Mehrotra C. What do binge drinkers drink? Implications for alcohol control policy. Am J Prev Med 2007;33(3):188–93.
11. Clapp JD, Shillington AM. Environmental predictors of heavy episodic drinking. Am J Drug Alcohol Abuse 2001;27(2):301–13.
12. Rogers JD, Greenfield TK. Beer drinking accounts for most of the hazardous alcohol consumption reported in the U.S. J Stud Alcohol 1999;60(6):732–9.
13. Berger DE, Snortum JR. Alcoholic beverage preferences of drinking–driving violators. J Stud Alcohol 1985;46(3):232–9.
14. Klein H, Pittman DJ. Drinker prototypes in American society. J Subst Abuse 1990;2(3):299–316.
15. Jensen MK, Andersen AT, Sorensen TI, Becker U, Thorsen T, Gronbaek M. Alcoholic beverage preference and risk of becoming a heavy drinker. Epidemiol Community Health 2002;13(2):127–32.
16. Klatsky AL, Armstrong MA, Kipp H. Correlates of alcoholic beverage preference: traits of persons who choose wine, liquor or beer. Br J Addict 1990;85(10):1279–89.
17. Smart RG. Behavioral and social consequences related to the consumption of different beverage types. J Stud Alcohol 1996;57(1):77–84.
18. Snortum JR, Kremer LK, Berger DE. Alcoholic beverage preference as a public statement: self-concept and social image of college drinkers. J Stud Alcohol 1987;48(3):243–51.
19. Schmid H, TerBogt T, Godeau E, Hublet A, Dias SF, Fotiou A. Drunkenness among young people: a cross-national comparison. J Stud Alcohol 2003;64(5):650–1.
20. Roeber J, Green DL, Meurer KM, et al. Types of alcoholic beverages usually consumed by students in 9th–12th grades—four states, 2005. MMWR Morb Mortal Wkly Rep 2007;56(29):737–40.
21. Cremeens JL, Miller JW, Nelson DE, Brewer RD. Assessment of source and type of alcohol consumed by high school students: analyses from four states. J Addict Med 2009;3(4):204–10.
22. Brener ND, Kann L, Kinchen SA, et al. Methodology of the youth risk behavior surveillance system. MMWR Recomm Rep 2004;53(RR-12):1–13.
23. van den Bulck J, Beullens K. Television and music video exposure and adolescent alcohol use while going out. Alcohol Alcohol 2005;40(3):249–53.
24. Atkin CK. Effects of televised alcohol messages on teenage drinking patterns. J Adolesc Health Care 1990;11(1):10–24.
25. Tucker LA. Television's role regarding alcohol use among teenagers. Adolescence 1985;20(79):593–8.
26. Woodruff RS. A simple method for approximating the variance of a complicated estimate. J Am Stat Assoc 1971;66(334):411–4.
27. SAS Institute Inc. Variance estimation. SAS/STAT(R) 9.22 user's guide. Cary NC: SAS Institute Inc., 2010. support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_surveyfreq_a0000000212.htm.
28. U.S. Census Bureau. 2007 population estimates. T6-2007: Sex by age. Washington DC: U.S. Census Bureau, Population Estimates Program. factfinder.census.gov.
29. The Center on Alcohol Marketing and Youth. Youth exposure to alcohol advertising on television and in national magazines, 2001 to 2006. Washington DC: Center on Alcohol Marketing and Youth, Georgetown University, 2007.
30. Jernigan D, Ostroff J, Ross C, et al. Youth exposure to alcohol advertising in magazines—U.S., 2001–2005. MMWR Morb Mortal Wkly Rep 2007;56(30):763–7.
31. Johnston LD, O'Malley PM, Bachman JG, et al. Monitoring the Future national survey results on drug use, 1975–2008. Volume I: secondary school students. Bethesda MD: National Institute on Drug Abuse, 2009. NIH Publication No. 09-7402.
32. The Center on Alcohol Marketing and Youth. Youth exposure to alcohol advertising on television, 2001 to 2007. Washington DC: Center on Alcohol Marketing and Youth, Georgetown University, 2008.

Supplementary data

A pubcast created by the authors of this paper can be viewed at http://www.ajpm-online.net/content/video_pubcasts_collection.

ARTICLE

Gender-Based Disparities in Infant and Child Mortality Based on Maternal Exposure to Spousal Violence

The Heavy Burden Borne by Indian Girls

Jay G. Silverman, PhD; Michele R. Decker, ScD, MPH; Debbie M. Cheng, PhD; Kathleen Wirth, ScM; Niranjan Saggurti, PhD; Heather L. McCauley, MS; Kathryn L. Falb, MPH; Balaiah Donta, PhD; Anita Raj, PhD

Objectives: To examine associations between intimate partner violence (IPV) against Indian women and risk of death among their infants and children, as well as related gender-based disparities.

Design: Analyses of nationally representative data to estimate adjusted hazard ratios (aHRs) and attributable risks for infant and child mortality based on child gender and on IPV against mothers.

Setting: India.

Participants: Women aged 15 to 49 years (n = 59 467) across all 29 Indian states participating in the Indian National Family Health Survey 3 provided information about 158 439 births and about infant and child mortality occurring during the 20 years before the survey.

Main Outcome Measures: Maternal IPV and infant and child (<5 years) mortality among boy vs girl children.

Results: Infant mortality was greater among infants whose mothers experienced IPV (79.2 of 1000 births) vs those whose mothers did not experience IPV (59.1 of 1000 births) (aHR, 1.09; 95% confidence interval [CI], 1.03-1.15); this effect was significant only for girls (1.15; 1.07-1.24; for boys, 1.04; 0.97-1.11). Child mortality was also greater among children whose mothers experienced IPV (103.6 of 1000 births) vs those whose mothers did not experience IPV (74.8 per 1000 births) (aHR, 1.10; 95% CI, 1.05-1.15); again, this effect was significant only for girls (1.14; 1.07-1.21; for boys, 1.05; 0.99-1.12). An estimated 58 021 infant girl deaths and 89 264 girl child deaths were related to spousal violence against wives annually, or approximately 1.2 million female infant deaths and 1.8 million girl deaths in India between December 1985 and August 2005.

Conclusion: Intimate partner violence against women should be considered an urgent priority within programs and policies aimed at maximizing survival of children in India, particularly those attempting to increase the survival of girls 5 years and younger.

Arch Pediatr Adolesc Med. 2011;165(1):22-27

Author Affiliations: Department Society, of Human Development and Health, Harvard School of Public Health (Drs Silverman and Decker and Mss Wirth, McCauley, and Falb) and Boston University School of Public Health (Drs Cheng and Raj), Boston, Massachusetts; and Population Council, New Delhi (Dr Saggurti), and National Institute for Research in Reproductive Health, Indian Council for Medical Research, Mumbai (Dr Donta), India.

APPROXIMATELY 10 MILLION children die across the globe each year before their fifth birthday.[1] One in 5 of these children (2.1 million) die in India.[1,2] Child mortality is a stubborn problem across South Asia. With more than 8% of live births estimated to end in death before age 5 years,[1,3] the region is failing to meet Millennium Development Goal 4 (reducing child mortality by two-thirds from 1990 to 2015).[1] As other regions in Asia have made greater strides toward this goal, South Asia's fraction of infant deaths has steadily risen, now estimated at approximately 80%.[1] With India alone contributing more than one-quarter of all births worldwide each year,[1] reductions in national infant and child mortality, even of a modest magnitude, are critical.

Multiple infant and child mortality risk factors have been well documented (eg, malnutrition, low birth weight, and infectious disease)[4-7]; however, less attention has been paid to the potential role of violence against mothers of infants and children. Violence from husbands has been implicated in child malnutrition,[8,9] low birth weight,[10,11] and infectious disease,[12] as well as direct violence against a child,[13] suggesting its role in mortality outcomes. Investigation into the role of intimate partner violence (IPV) in mortality outcomes is indicated in India given its high prevalence of spousal violence against wives (>1 in 3),[14,15] coupled with high rates of infant and child mortality.[1]

Gender disparity constitutes a critical yet understudied issue regarding Indian infant and child mortality. Differential mal-

treatment of girls vs boys has been discussed extensively,[1,7,16-18] and millions of girls have been described as "missing" in the population.[19] The high ratio of male births to female births in India[20] is thought to reflect sex-selective abortion and undocumented female infanticide and other female infant death.[5,16,19,21,22] These gender-based disparities extend to child mortality in India[23]; an estimated 70 of 1000 boys born will die before age 5 years, while this estimate is 13% higher for girls at 79 of 1000 born.[1] Notably, the pattern in India is the reverse of that found globally (ie, child survival is greater for girls in most other nations[24,25]).

To date, there has been little empirical assessment as to whether and to what extent spousal violence affects the likelihood of infant and child survival. Four district or community-based studies have explored this question to date, 2 studies[26,27] in rural districts of the Indian states of Uttar Pradesh and Tamil Nadu, 1 study[28] in a rural district in Bangladesh, and 1 case-control study[29] in Leon, Nicaragua. Results of 3 of these studies[26,27,29] using mothers as the unit of analysis indicate that a woman's history of violence from a male partner increases the likelihood of infant or child mortality. The single study[28] of community-based data that assessed the association of violence against mothers with the risk of child mortality at the child level rather than at the maternal level found no overall elevated risk of mortality conferred by such violence. A 2009 study[30] examined this relationship using Indian data and found evidence of an association, but the findings cannot be considered nationally representative because sample weights were not used; furthermore, the analyses were not stratified by child gender.

To advance the state of knowledge, this study used a large national data set weighted to provide a nationally representative sample of mothers' reports of infant and child deaths in India collected via the Indian National Family Health Survey 3 (NFHS-3) between November 2005 and August 2006. Analyses were conducted to (1) assess whether violence against mothers was related to elevated rates of infant and child (aged <5 years) mortality, (2) evaluate if such rates differed for girls vs boys, and (3) explore the extent to which recent national census statistics reflect such mortality rates and related gender-based disparities.

METHODS

DESIGN, SETTING, AND SAMPLE

From November 2005 to August 2006, the NFHS-3 was conducted in all 29 Indian states by the International Institute for Population Sciences and Macro International.[15] The NFHS, also referred to as the Demographic Health Survey in other national contexts, is conducted regularly in many developing countries to obtain population-based estimates of major health threats. This surveillance involves confidential questionnaires administered verbally in private locations within sampled households; surveys were bilingual within each state, with questions available in English and in the principal language of that Indian state. The nationally representative household-based sample for the NFHS-3 was created via a stratified multistage cluster strategy. Within each state, 2-stage (rural areas) and 3-stage (urban areas) procedures identified 3850 primary sampling units comprising 1 or more villages in rural areas and census enumeration blocks within wards in urban areas; primary sampling unit selection probability was proportional to population size. Within each primary sampling unit, household enumeration generated the sampling frame for systematic selection of households. Trained research assistants conducted household-based recruitment and obtained written informed consent immediately before survey data collection. Further details concerning the NFHS-3 procedures have been published previously.[15]

These procedures identified 131 596 eligible women in India aged 15 to 49 years, of whom 124 385 completed the survey (response rate, 94.5%).[15] Although the overall sampling strategy allowed for multiple female participants per household, a separate systematic procedure selected a single female participant to complete the IPV assessment. This assessment was conducted by female interviewers only when the privacy of the respondent was guaranteed.[31] All interview responses were kept strictly confidential and were not shared with other members of the household. This procedure was designed to maximize participant confidentiality in responding to sensitive items concerning violence victimization and to prevent risk to any individual based on subsequent discussion of the assessment among participating household members. Of 124 385 female survey participants, 84 268 (67.7%) were selected for the IPV module, and 83 703 of these (99.3%) completed the module. Each participant was asked to enumerate the births of their children, including gender, current age, and age at death if applicable. Among 83 703 women who completed the IPV assessment, 63 356 had given birth; these participants provided birth information about 187 351 children, which serve as the unit of analysis for the present investigation. The analytic sample was further restricted to births within the past 20 years (n=159 053) to maximize inferences to the current population while retaining statistical power. Six hundred fourteen births with incomplete data concerning violence exposure were excluded, resulting in a final sample size of 59 467 mothers and 158 439 births.

MEASURES

Maternal demographics, including age, parity, and education, were assessed via single self-reported items. A relative index of household wealth was calculated based on interviewer-observed assets (eg, ownership of consumer items), with resulting scores divided into quintiles (1 indicated the lowest level of household wealth; 5, the highest level of household wealth). The primary exposure, IPV, was assessed via self-report in accord with World Health Organization recommendations[32,33] and was based on the Revised Conflict Tactics Scale.[34] Lifetime IPV victimization was indicated by a positive answer to any of 8 items pertaining to whether their current husband had performed the following: "push you, shake you, or throw something at you"; "slap you"; "punch you with a fist or something harmful"; "kick, drag, or beat you up"; "try to choke or burn you on purpose"; "threaten or attack you with a knife, gun, or any other weapon"; "physically force you to have sexual intercourse with him even when you did not want to"; or "force you to perform any sexual acts that you did not want to." Items demonstrated adequate internal consistency reliability; Cronbach α was .76 for the sample. Infant and child mortality outcomes were assessed via participant enumeration of each live birth, including date of birth, current age, and age at death if applicable; consistent with international standards, infant mortality was defined as death before age 12 months, and child mortality was defined as death before age 60 months.

All data collection procedures were approved by the ORC Macro International Institutional Review Board. The Harvard School of Public Health Human Subjects Committee deemed secondary analyses exempt given the anonymous nature of the data.

Table 1. Demographic Characteristics of 59 467 Indian Women Giving Birth in the 20 Years (1985-2005) Before the Survey[a]

Variable	Total Sample, % (95% CI)	IPV Exposed, % (95% CI)[b]
Overall	**100**	34.1 (33.6-34.6)
Maternal age, y		
<25	18.5 (18.1-18.9)	34.4 (33.3-35.6)
25-29	21.3 (20.9-21.7)	34.9 (33.2-35.0)
30-34	20.6 (20.2-20.9)	34.0 (33.1-34.9)
35-39	18.8 (18.4-19.2)	34.8 (33.8-35.9)
40-44	13.6 (13.3-14.0)	33.0 (31.7-34.4)
≥45	7.2 (6.9-7.5)	33.7 (31.8-35.6)
Highest level of education achieved		
No education	43.9 (43.4-44.4)	43.6 (42.9-44.4)
Primary	16.0 (15.6-16.3)	37.5 (36.3-38.7)
Secondary	33.8 (33.4-34.3)	24.6 (23.9-25.3)
Higher	6.3 (6.1-6.5)	10.0 (9.0-11.1)
Household wealth index		
Poorest	16.3 (15.9-16.6)	48.3 (47.1-49.5)
Poor	18.5 (18.4-18.9)	43.8 (42.7-45.0)
Middle	21.1 (20.7-21.5)	37.5 (36.5-38.6)
Richer	21.5 (21.1-21.9)	30.4 (29.5-31.4)
Richest	22.7 (22.3-23.1)	16.3 (15.5-17.0)
Dwelling		
Urban	31.9 (31.5-32.3)	28.2 (27.5-29.0)
Rural	68.1 (67.7-68.5)	36.8 (36.3-37.4)

Abbreviations: CI, confidence interval; IPV, intimate partner violence.
[a]The mean (SE) parity was 3.10 (0.01) for the total sample and 3.50 (0.02) for the IPV exposed ($P < .001$, Pearson product moment χ^2 test).
[b]Row percentages. *P* values (Pearson product moment χ^2 test) are .41 for maternal age and <.001 for highest level of education achieved, household wealth index, and dwelling.

STATISTICAL ANALYSIS

Prevalence estimates of maternal IPV victimization were calculated for the total sample of mothers (n=59 467). Differences in IPV prevalence based on demographics were assessed using the Wald χ^2 analysis; significance for all analyses was set at $P < .05$.

Infant and child mortality rates per 1000 live births were calculated for the total sample of births (n=158 439) and by gender overall and stratified based on maternal exposure to IPV. The vital statistics method was chosen for the analyses (ie, the proportion of deaths per live births) rather than derivation of mortality via calculation of component death rates, as is practiced by Macro International. Because both methods of mortality rate calculation are limited by several potentially false assumptions (eg, accurate reporting of exact month-level date of a child's death and linear changes in mortality), the vital statistics method was chosen because of its ability to describe the actual number of deaths per 1000 births. Time-to-event analysis using Cox proportional hazards models was conducted to evaluate the association of maternal exposure to IPV with infant and child mortality using person-months as the unit of time. Models were subsequently adjusted for maternal demographic covariates (maternal age, educational status, parity, rural dwelling, and household wealth index). Effect modification by gender was assessed by creating an interaction term of maternal exposure to IPV and child gender and by including it in the multivariate-adjusted Cox proportional hazards model. The interaction term was evaluated using the Wald χ^2 analysis. Finally, adjusted hazard ratios (aHRs) were used to calculate the population-attributable fraction (PAF) to estimate the fractions of all infant and child mortality cases that would not have occurred in the absence of maternal violence (overall and by child gender). The PAF was calculated using the following computation: $p_c([RR-1]/RR)$, where p_c represents the IPV prevalence among the cases, and *RR* indicates risk ratio.[35,36] To estimate the effect of IPV on infant and child mortality during the 20-year period (December 1985 to August 2005), the PAF was then multiplied by the corresponding annual numbers of infant and child births and deaths by gender that occurred during that period.[15,20,37-39]

All analyses were performed using commercially available statistical software (SAS version 10.0; SAS Institute, Cary, North Carolina); survey analysis procedures were used to accommodate the stratified cluster sampling design of the NFHS-3 and the potential for nonindependence of responses within primary sampling units. Analyses were weighted for nonresponse using the nationally representative women's IPV module weights.

RESULTS

IPV PREVALENCE AND ASSOCIATED DEMOGRAPHICS

More than 1 in 3 married Indian women who had given birth in the past 20 years (34.1%) reported having experienced IPV (**Table 1**). This estimate is slightly lower than the 40% prevalence reported for the overall NFHS-3 sample.[15]

INFANT AND CHILD MORTALITY BASED ON IPV EXPOSURE AND CHILD GENDER

Infant mortality was significantly greater among births to mothers experiencing spousal violence (79.2 of 1000 births) vs those who did not experience spousal violence (59.1 of 1000 births) (significance was based on nonoverlapping 95% confidence intervals [CIs]). The aHR for this effect (1.09; 95% CI, 1.03-1.15) indicates an estimated 9% increase in the risk of infant mortality based on maternal violence exposure (**Table 2**).

Analyses stratified by child gender demonstrated that this association was pronounced and significant only for girl infants. Among infant girls, maternal exposure to IPV conferred elevated risk for infant mortality (HR, 1.15; 95% CI, 1.07-1.24), while for infant boys it was nonsignificant (1.04; 0.97-1.11) .

Similarly, children younger than 5 years were significantly more likely to die if their mothers experienced spousal violence (103.6 vs 74.8 deaths per 1000 births) vs those whose mothers did not experience such violence (74.8 of 1000 births). In adjusted Cox proportional hazards regression models, maternal exposure to IPV significantly elevated the risk of child mortality by approximately 10% (aHR, 1.10; 95% CI, 1.05-1.15). Similar to results for infant mortality, analyses stratified by child gender demonstrated that the elevated mortality was borne primarily by girls, with their risk of death by age 5 years increased by 14% based on maternal exposure to IPV (aHR, 1.14; 95% CI, 1.07-1.21). Among boys, IPV was not associated with a significant increase in mortality risk (aHR, 1.05; 95% CI, 0.99-1.12).

In analyses containing an interaction term for child gender and maternal exposure to IPV, the Wald χ^2 analy-

Table 2. Infant and Child (<5 Years) Mortality by Maternal Exposure to Intimate Partner Violence (IPV) Among 158 439 Births (74 054 Female and 84 385 Male)

Variable	Mortality (95% CI) per 1000 Births[a]		Hazard Ratio (95% CI)	Adjusted Hazard Ratio (95% CI)[b]	Population-Attributable Fraction (95% CI) for IPV
	Among IPV Exposed	Among IPV Unexposed			
Infant mortality					
Overall	79.2 (76.3-82.2)	59.1 (57.0-61.1)	1.34 (1.27-1.41)	1.09 (1.03-1.15)	0.0366 (0.0122-0.0604)
Female	76.9 (72.9-81.0)	54.6 (51.8-57.3)	1.42 (1.32-1.53)	1.15 (1.07-1.24)	0.0607 (0.0243-0.0958)
Male	81.4 (77.6-85.3)	63.3 (60.5-66.1)	1.28 (1.20-1.38)	1.04 (0.97-1.11)	. . .
Child (aged <5 y) mortality					
Overall	103.6 (100.3-106.9)	74.8 (72.5-77.2)	1.39 (1.33-1.46)	1.10 (1.05-1.15)	0.0398 (0.0182-0.0610)
Female	105.4 (100.8-110.0)	73.0 (69.9-76.1)	1.46 (1.37-1.55)	1.14 (1.07-1.21)	0.0577 (0.0266-0.0879)
Male	101.9 (97.7-106.2)	76.5 (73.6-79.6)	1.33 (1.26-1.41)	1.05 (0.99-1.12)	. . .

Abbreviations: CI, confidence interval; ellipses, not applicable.
[a] Calculated as the proportion of live births resulting in death multiplied by 1000.
[b] Adjusted for maternal age, education, parity, rural (vs urban) residence, and household wealth index.

sis indicated that the observed differences between girl and boy mortality were statistically significant. *P* values for the interaction terms were .02 for child mortality and .03 for infant mortality.

The PAF was calculated to estimate the potential effect of reducing IPV on rates of infant and child mortality (ie, the fraction of deaths that may be attributable to the direct or indirect effects of IPV); the PAF is estimated only for those effects found to be significant based on hazard ratios. The PAFs were 3.66% (95% CI, 1.22%-6.04%) for overall infant mortality and 6.07% (2.43%-9.58%) for infant girl mortality (Table 2). The PAFs were 3.98% (95% CI, 1.82%-6.10%) for overall child mortality and 5.77% (2.66%-8.79%) for female child mortality. These findings suggest that approximately 1 in 15 deaths of female infants and 1 in 16 deaths of female children could be prevented by eliminating IPV.

To estimate the numbers of infant deaths overall and infant girl deaths attributable to IPV each year across the period studied, as well as across the entire period, PAFs were multiplied by the number of annual infant deaths and infant girl deaths per UNICEF (United Nations International Children's Emergency Fund)[37] and World Health Organization[39] statistics for the midpoint of this interval (ie, 1995). Infant mortality statistics from 1995 closely mirror the means of those derived for 1990 and 2000 and for 1985 and 2005 by the Indian government,[31] indicating the validity of using 1995 mortality statistics in calculating the mean annual mortality (further details about the computations are available from the corresponding author). On the basis of these calculations,[38-40] spousal violence in India against wives may account for 72 617 infant deaths each year across the 20-year period from 1985 to 2005, with girls representing 58 021 of these annual deaths. Considered over the 20-year period under study, spousal violence against mothers is associated with an estimated 1 160 440 girl infant deaths. Using the same method of computation,[38-40] spousal violence in India may be related to approximately 119 480 child deaths each year from 1985 to 2005, with girls representing 89 264 of these annual deaths. Therefore, across the 2 decades studied, violence against mothers is associated with an estimated 1 785 280 female child deaths.

COMMENT

Infants and young children in India were found to suffer significantly greater risk of death in families in which mothers had experienced spousal violence from their husbands. Furthermore, the effect of such gender-based violence was profoundly gendered; infant girls and children bear a far greater share of the mortality burden associated with IPV. In contrast, IPV was not significantly associated with infant boys and child mortality in adjusted analyses. Even after considering the birth of fewer girls than boys in India, deaths of infant girls and young girls accounted for an estimated 80% of all infant deaths and 75% of all child deaths related to IPV, translating to approximately 58 021 infant girls and 89 264 young girls dying each year from 1985 to 2005, or 1.2 million infant girl deaths and 1.8 million girl deaths in India across the 20-year period studied herein.

Violence against mothers may contribute to infant and child mortality by various mechanisms. Evidence of a significant gender differential in this pattern suggests that these mechanisms disproportionately affect girls, likely reflecting the perceived relative lower value of girls vs boys in many Indian families and communities. Moreover, violence against wives may well be a marker for multiple other forms of gender-based maltreatment and neglect of girls (eg, provision of less food, reduced attention to infection prevention, and decreased investment in care for illness).

Because violence against wives is considered an expression of men's sense of entitlement to use violence to control women (including women's care of children),[33,41] it is reasonable to assume that this same belief system informs such men's treatment of their girls.[33,41] Men who perpetrate IPV may well see girls as less deserving of care, leading to inadequate provision of food, hygiene, and other preventive measures, as well as reluctance to invest in the treatment of girls in case of illness; thus, such men ensure that the resources of the family are directed to adult men and sons.[42] In the most extreme forms, violence or mistreatment toward infant girls may take the form of female infanticide.[33,42] Evidence of this gender-based disparity was recently described among a large Indian sample

in which socioeconomic status was associated with the provision of health care only for boys[43]; such effects may be magnified in abusive homes.

Moreover, although child abuse was not assessed herein, it may be that men who are violent against their wives also exhibit violence and mistreatment toward their children.[13] The relative low worth and social and economic costs traditionally associated with girls[44] may make such abuse more likely to extend to girls 5 years and younger.

Abused women may be less able to care for their children based on the emotional and physical sequelae of the violence they have experienced,[45] with such violence being particularly severe following the birth of a girl for the reasons aforementioned.[16,46] The incapacitation of a mother may affect male children to a lesser extent, as family members may be more likely to assist with boys in such cases.[28]

Violence has been found to affect the pregnancy-related health of women and to increase the likelihood of having a low-birth-weight infant.[10,11] However, this explanation does not account for the observed gender disparity in the association of IPV and mortality, as the sex of the child will most likely remain unknown until birth.

Several important limitations of the present study design should be considered in reviewing these findings. The cross-sectional nature of the investigation does not allow for conclusions regarding temporality. In other words, infant and child deaths may precede violence against Indian wives. However, it is unclear why violence would be more likely in cases of the death of a girl 5 years or younger. As discussed, IPV may well be a marker for other gender-related conditions that affect female mortality; to clarify the present findings, longitudinal study regarding a broad spectrum of gender-based maltreatment of women and its potential effects on female child mortality is necessary. Several potential mechanisms underpinning our results were unable to be formally evaluated; data were not collected about the cause of death, abuse of children, or nutritional and health status among children who had died. Such data are needed to better understand the associations identified. Based on the nature of our research question, the present sample was limited to women who were systematically selected for and completed the IPV assessment module. Despite the high participation rate in the IPV module (99.3%), it is possible that this subsample of female participants and, most important, the births they described were in some way nonrepresentative of the larger survey sample and the population. Finally, although child deaths are critical events in life, the reliability of women's recall of such events across the preceding 20 years is unknown.

In summary, the magnitude of the association of IPV perpetration by Indian men with child mortality, as well as the numbers of infants and children who may die based on this modifiable factor, is great; so too is the disproportionate burden borne by young Indian girls. A clear implication of the present findings is the urgent need for IPV against women to be considered an urgent priority within programs and policies aimed at maximizing survival of children in India, particularly those attempting to increase the survival of girls. Violence against mothers[2,29] and the associated gender-based mistreatment of female infants and children[2,23,24] may represent major barriers preventing India from reaching the Millennium Development Goal 4 of a two-thirds reduction in child mortality from 1990 levels by 2015.[1] Regardless of such targets and deadlines, violence against wives in India must be vigorously challenged. Even a modest reduction in the prevalence of IPV may prevent the deaths of tens of thousands of Indian infant girls and children. Such progress may be accomplished by interventions that address gender norms within the context of maternal, neonatal, and child health (eg, antenatal programs for men who are identified as perpetrating IPV).

Our finding that approximately 1 in 15 female infant deaths and 1 in 16 fermale child deaths may be prevented by eliminating IPV should lead to greater investment in changing discriminatory gender norms among Indian boys and men. Norms should be altered that lead to a broad range of physical and social hazards for women and girls, including gender-based violence and other maltreatment of wives, as well as the associated gendered increase in infant and child mortality suffered by girls born into such families.

Accepted for Publication: June 2, 2010.
Correspondence: Jay G. Silverman, PhD, Department of Human Development and Health, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115 (jsilverm@hsph.harvard.edu).
Author Contributions: Dr Silverman had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Silverman, Decker, Saggurti, Donta, and Raj. *Acquisition of data:* Decker and Wirth. *Analysis and interpretation of data:* Silverman, Decker, Cheng, Wirth, McCauley, and Falb. *Drafting of the manuscript:* Silverman, Decker, Cheng, and Wirth. *Critical revision of the manuscript for important intellectual content:* Silverman, Decker, Saggurti, McCauley, Falb, Donta, and Raj. *Statistical analysis:* Silverman, Decker, Cheng, Wirth, McCauley, and Raj. *Obtained funding:* Silverman and Donta. *Administrative, technical, and material support:* Silverman, Cheng, Saggurti, Falb, Donta, and Raj. *Study supervision:* Decker.
Financial Disclosure: None reported.
Funding/Support: This study was supported by grant 1R03HD055120-01 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and by the Indian Council for Medical Research (Drs Silverman and Donta).

REFERENCES

1. United Nations International Children's Emergency Fund. *The State of Asia-Pacific's Children 2008: Child Survival*. New York, NY: UNICEF; 2008.
2. van der Klaauw B, Wang L. *Child Mortality in Rural India*. Washington, DC: World Bank; 2004.
3. World Health Organization. *The World Health Report 2005: Make Every Mother and Child Count*. Geneva, Switzerland: World Health Organization; 2005.
4. United Nations International Children's Emergency Fund. *The State of the World's Children 2009*. New York, NY: UNICEF; 2008.
5. Lawn JE, Cousens S, Zupan J; Lancet Neonatal Survival Steering Team. 4 million neonatal deaths: when? where? why? *Lancet*. 2005;365(9462):891-900.
6. Black RE, Morris SS, Bryce J. Where and why are 10 million children dying every year? *Lancet*. 2003;361(9376):2226-2234.
7. Khanna R, Kumar A, Vaghela JF, Sreenivas V, Puliyel JM. Community based retrospective study of sex in infant mortality in India. *BMJ*. 2003;327(7407):126.

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC165706/?tool=pubmed. Accessed October 5, 2010.
8. Ackerson LK, Subramanian SV. Domestic violence and chronic malnutrition among women and children in India [published correction appears in *Am J Epidemiol*. 2009;170(2):268]. *Am J Epidemiol*. 2008;167(10):1188-1196.
9. Hasselmann MH, Reichenheim ME. Parental violence and the occurrence of severe and acute malnutrition in childhood. *Paediatr Perinat Epidemiol*. 2006; 20(4):299-311.
10. Valladares E, Ellsberg M, Peña R, Högberg U, Persson LA. Physical partner abuse during pregnancy: a risk factor for low birth weight in Nicaragua. *Obstet Gynecol*. 2002;100(4):700-705.
11. Silverman JG, Decker MR, Reed E, Raj A. Intimate partner violence victimization prior to and during pregnancy among women residing in 26 U.S. states: associations with maternal and neonatal health. *Am J Obstet Gynecol*. 2006;195 (1):140-148.
12. United Nations Population Fund. *State of World Population 2005: The Promise of Equality, Gender Equity, Reproductive Health and the Millennium Development Goals*. Geneva, Switzerland: UNFPA; 2005.
13. Edleson JL. The overlap between child maltreatment and woman battering. *Violence Against Women*. 1999;5(2):134-154.
14. Silverman JG, Decker MR, Saggurti N, Balaiah D, Raj A. Intimate partner violence and HIV infection among married Indian women. *JAMA*. 2008;300(6): 703-710.
15. International Institute for Population Sciences. *National Family Health Survey (NFHS-3) 2005-06: India*. Vol 1. Mumbai, India: IIPS; 2007.
16. Sheth SS. Missing female births in India [comment]. *Lancet*. 2006;367(9506):185-186.
17. Sharma DC. Widespread concern over India's missing girls: selective abortion and female infanticide cause girl-to-boy ratios to plummet. *Lancet*. 2003;362 (9395):1553.
18. Ahmed S, Sobhan F, Islam A, Barkat-e-Khuda. Neonatal morbidity and care-seeking behaviour in rural Bangladesh. *J Trop Pediatr*. 2001;47(2):98-105.
19. Jha P, Kumar R, Vasa P, Dhingra N, Thiruchelvam D, Moineddin R. Low female [published correction appears in *Lancet*. 2006;367(9524):1730]-to-male [corrected] sex ratio of children born in India: national survey of 1.1 million households. *Lancet*. 2006;367(9506):211-218.
20. United Nations Department of Economic and Social Affairs. *World Population Prospects: The 2006 Revision*. New York, NY: United Nations; 2007.
21. Sharma BR, Gupta N, Relhan N. Misuse of prenatal diagnostic technology for sex-selected abortions and its consequences in India. *Public Health*. 2007; 121(11):854-860.
22. Arnold F, Kishor S, Roy TK. Sex-selective abortions in India. *Popul Dev Rev*. 2002; 28(4):759-785.
23. Claeson M, Bos ER, Mawji T, Pathmanathan I. Reducing child mortality in India in the new millennium. *Bull World Health Organ*. 2000;78(10):1192-1199.
24. Watts C, Zimmerman C. Violence against women: global scope and magnitude. *Lancet*. 2002;359(9313):1232-1237.
25. Sen A. Missing women. *BMJ*. 1992;304(6827):587-588.
26. Ahmed S, Koenig MA, Stephenson R. Effects of domestic violence on perinatal and early-childhood mortality: evidence from north India. *Am J Public Health*. 2006;96(8):1423-1428.
27. Jejeebhoy SJ. Associations between wife-beating and fetal and infant death: impressions from a survey in rural India. *Stud Fam Plann*. 1998;29(3):300-308.
28. Åsling-Monemi K, Tabassum Naved R, Persson LA. Violence against women and the risk of under-five mortality: analysis of community-based data from rural Bangladesh. *Acta Paediatr*. 2008;97(2):226-232.
29. Åsling-Monemi K, Peña R, Ellsberg MC, Persson LÃ. Violence against women increases the risk of infant and child mortality: a case-referent study in Nicaragua. *Bull World Health Organ*. 2003;81(1):10-16.
30. Ackerson LK, Subramanian SV. Intimate partner violence and death among infants and children in India. *Pediatrics*. 2009;124(5):e878-e889. doi:10.1542/peds .2009-0524. Accessed December 22, 2009.
31. International Institute for Population Sciences. National Family Health Survey, India; 2007. http://www.nfhsindia.org/nfhs3.html. Accessed November 1, 2010.
32. World Health Organization. *Preventing Violence: A Guide to Implementing the Recommendations of the World Report on Violence and Health*. Geneva, Switzerland: WHO; 2004.
33. World Health Organization. *World Report on Violence and Health*. Geneva, Switzerland: WHO; 2002.
34. Straus MA, Hamby SL, Boney-McCoy S, Sugarman DB. The Revised Conflict Tactics Scale (CTS2): development and preliminary psychometric data. *J Fam Issues*. 1996;17(3):283-316.
35. Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am J Epidemiol*. 1974;99(5):325-332.
36. Rothman KJ, Greenland S, eds. *Modern Epidemiology*. 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 1998.
37. United Nations International Children's Emergency Fund. Statistics: information by country: India; 2009. http://www.unicef.org/infobycountry/india_statistics .html. Accessed July 27, 2009.
38. United Nations International Children's Emergency Fund. *The State of the World's Children 1997*. New York, NY: UNICEF; 1997.
39. World Health Organization Statistical Information System. Statistics. 2010. http: //apps.who.int/whosis/data/Search.jsp. Accessed February 9, 2010.
40. Kulkarni PM. *Estimation of Missing Girls at Birth and Juvenile Ages in India*. New Delhi, India: United Nations Population Fund; 2007.
41. Ahmed-Ghosh H. Chattels of society: domestic violence in India. *Violence Against Women*. 2004;10(1):94-118.
42. Fikree FF, Pasha O. Role of gender in health disparity: the South Asian context. *BMJ*. 2004;328(7443):823-826.
43. Bhan G, Bhandari N, Taneja S, Mazumder S, Bahl R; Zinc Study Group. The effect of maternal education on gender bias in care-seeking for common childhood illnesses. *Soc Sci Med*. 2005;60(4):715-724.
44. Arnold F, Choe MK, Roy TK. Son preference, the family-building process and child mortality in India. *Popul Studies*. 1998;52(3):301-315.
45. Levendosky AA, Graham-Bermann SA. Parenting in battered women: the effects of domestic violence on women and their children. *J Fam Violence*. 2001;16 (2):171-192.
46. Das Gupta M. Asia's "missing women": a new look at the data. *Popul Dev Rev*. 2005;31(3):529-535.

You can't stay in your corner of the Forest waiting for others to come to you. You have to go to them sometimes.
—Winnie the Pooh

Sex Trafficking and Initiation-Related Violence, Alcohol Use, and HIV Risk Among HIV-Infected Female Sex Workers in Mumbai, India

Jay G. Silverman,[1] Anita Raj,[1] Debbie M. Cheng,[2] Michele R. Decker,[5] Sharon Coleman,[3] Carly Bridden,[4] Manoj Pardeshi,[6] Niranjan Saggurti,[7] and Jeffrey H. Samet[4]

[1]Division of Global Public Health, Department of Medicine, University of California at San Diego School of Medicine, La Jolla, [2]Department of Biostatistics, [3]Data Coordinating Center, Boston University School of Public Health, and [4]Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine/Boston Medical Center, Massachusetts, [5]Department of Population, Family and Reproductive Health, Johns Hopkins School of Public Health, Baltimore, Maryland; [6]Network of Maharashtra by People Living with HIV & AIDS, (NMP+), Pune, and [7]Population Council, New Delhi, India

Female sex workers (FSWs) are the group at greatest risk for human immunodeficiency virus (HIV) infection in India. Women and girls trafficked (ie, forced or coerced) into sex work are thought to be at even greater risk because of high exposure to violence and unprotected sex, particularly during the early months of sex work, that is, at initiation. Surveys were completed with HIV-infected FSWs (n = 211) recruited from an HIV-related service organization in Mumbai, India. Approximately 2 in 5 participants (41.7%) reported being forced or coerced into sex work. During the first month in sex work, such FSWs had higher odds of sexual violence (adjusted odds ratio [AOR], 3.1; 95% confidence interval [CI], 1.6–6.1), ≥7 clients per day (AOR, 3.3; 1.8–6.1), no use of condoms (AOR, 3.8, 2.1–7.1), and frequent alcohol use (AOR, 1.9; 1.0–3.4) than HIV-infected FSWs not entering involuntarily. Those trafficked into sex work were also at higher odds for alcohol use at first sex work episode (AOR, 2.2; 95% CI, 1.2–4.0). These results suggest that having been trafficked into sex work is prevalent among this population and that such FSWs may face high levels of sexual violence, alcohol use, and exposure to HIV infection in the first month of sex work. Findings call into question harm reduction approaches to HIV prevention that rely primarily on FSW autonomy.

There are currently 2.4 million persons in India living with human immunodeficiency virus (HIV) infection [1], with most having acquired the virus via heterosexual sex [2]. Female sex workers (FSWs) remain at greatest risk for HIV infection in India, with commercial sex involvement viewed as the primary means of transmission [1, 3]. The city of Mumbai is considered an epicenter for both sex work and HIV infection [4], with ≥10 000 FSWs and more HIV-related deaths than in any other location in India [5, 6].

Forced or coerced entry (ie, being "trafficked") into sex work is increasingly considered a marker for HIV risk among FSWs. However, although recent data from South Asia illustrate high HIV prevalence among former FSWs identified as having been trafficked [7, 8, 9], little is known of either the prevalence of this experience among general samples of FSWs, or how sexual risk may differ based on involuntary entry into sex work. A prominent explanation for why those trafficked into sex work are at greater risk for HIV infection is exposure to high levels of violence and related sexual risk during their involuntary initiation into prostitution [10]. Qualitative research among Indian and Nepali FSWs identified as trafficked indicates that violence in the period immediately after entry to sex work may involve high levels of sexual brutality, leading to vaginal injuries and significant blood loss, thus creating high

This article is part of a supplement entitled, "Sex Work in Asia: Health, Agency, and Sexuality," which was organized by the Harvard University Asia Center.

Correspondence: Jay G. Silverman, PhD, Division of Global Public Health, Department of Medicine, University of California at San Diego School of Medicine, 9500 Gilman Dr, MC 0507, La Jolla, CA 92093-0507 (jsilverman@ucsd.edu).

The Journal of Infectious Diseases 2011;204:S1229–34

0022-1899 (print)/1537-6613 (online)/2011/204S5-0006$14.00
DOI: 10.1093/infdis/jir540

vulnerability to sexually transmitted infection [8]. Two previous quantitative studies have compared sexual violence at entry and HIV risk based on having been trafficked; both demonstrated that sexual violence at entry is more common among those trafficked into sex work (in one case, those entering at <18 years of age were also coded as having been trafficked) than among FSWs not reporting this experience [7, 11]. However, no quantitative studies have examined the initial period of sex work (ie, after entry) to identify the prevalence and qualities of sexual risk and violence during this first month of "normal" commercial sex exposure, nor has the phenomenon been examined among FSWs infected with HIV. Alcohol has also been implicated in HIV risk among FSWs in India [12]. Almost half (44%) of FSWs in India report drinking alcohol [13], and a recent study of HIV-infected FSWs found that the majority of those who do use alcohol are heavy and/or dependent drinkers [12]. The aforementioned qualitative study of violence and HIV risk among trafficked South Asian women indicates that alcohol may be used to "initiate" women and girls into sex work involuntarily with heavy voluntary alcohol use continuing as a means for them to cope with their lack of autonomy [8]. To date, no quantitative studies have integrated assessments of forced or coerced entry with assessments of alcohol use. To advance the state of knowledge regarding the mechanisms that may link HIV infection and involuntary entry to sex work, the current study assessed the prevalence of trafficking among HIV-infected FSWs in Mumbai, India, and examined whether such history is associated with differences in history of sexual violence, sexual risk, and alcohol use in the initial 30 days of sex work. Such data are critical for prioritizing and developing HIV interventions designed to reach individuals in the early stages of sex work and illuminating the consequences of forced and coerced entry to sex work. Furthermore, because exposure to violence has been linked both to lower likelihood of adherence to antiretroviral therapy and to higher current sexual risk among HIV-infected FSWs, [8] the current findings may also inform secondary prevention efforts to reduce both mortality and the likelihood of HIV transmission among this vulnerable population.

METHODS

HIV infected FSWs were recruited from the ASHA Center, a community-based organization in Mumbai, managed and run by a group of FSWs who provide support and linkage to HIV-related healthcare. A total of 326 FSWs were contacted for study recruitment, of whom 246 (75%) responded positively. Of these, 216 individuals met the study's eligibility criteria: ≥18 years old, HIV infected, and reporting both sex trade involvement in the past year and penile-vaginal or anal sex in the past 30 days. HIV infection was confirmed by medical records brought by the participants. Of those eligible for the study, 97.7% (211/216) were willing to participate and complete survey interviews. Participants received a 45-minute interviewer-administered survey in Hindi assessing demographics, alcohol use, sex risk behaviors, and health status. Instruments were developed in English, translated into Hindi, and then reviewed by a study investigator fluent in both languages. Participants were given 100 rupees ($2.50) as compensation for their time in this study. Study procedures were reviewed and approved by the institutional review boards of Boston University Medical Campus, the Harvard School of Public Health, NMP+, and the Indian Council of Medical Research.

Measures

Demographic data were collected based on items modified or taken from the Indian Demographic and Health Survey [14]. The assessment of being trafficked into sex work, the main independent variable, was developed for this study based on the authors' previous research with sex-trafficked women and girls in South Asia [9, 15]; participants were asked "How did you start having sex for money?" and directed to select the answer they felt best described "how you got you into this work." (The wording of assessments reflects the back-translation from Hindi to English.) Options included (a) I decided myself that this was a good way for me to earn money; (b) someone told me that I should do this business, and I felt I had no choice but to enter; (c) someone forced me to come and do this business; (d) someone tricked me into coming to do this business; (e) I accepted a job doing some other kind of work and was made to have sex for money; and (f) someone used some other means besides trickery or force to make me enter. Participants selecting c, d, e, or f were classified as having been trafficked.

To assess experiences of sexual violence in the first month after entry, participants were asked "During the first month that you were in sex work, how often did anyone use violence or force to make you have sex or have certain types of sex with male clients?" Response options were "never," "rarely," "sometimes," or "very often." Those responding "sometimes" or "very often" were classified as experiencing violence in the first month of sex work. This same response set was used for determining condom nonuse during this same period with the question "During the first month in sex work, how often were condoms used when you were having sex with male clients?" This response pattern was again used to assess frequent alcohol use during the first month of sex work with the question "In the first month that you were in sex work, how often did you use alcohol when engaging in sex work?" Those who reported using alcohol "very often" were classified as frequently using alcohol during this period. Client volume during this initial period of exposure to sex work was assessed by asking "During the first month that you were in sex work, approximately how many male clients did you have on a working day?" Possible responses included 1–2, 3–6, and ≥7;

Table 1. Demographic Characteristics of HIV-Infected Female Sex Workers in Mumbai, India, and Demographic Differences Based on Experiences of Being Trafficked Into Sex Work

Demographic characteristic	Female sex workers, % (No.)			P
	Total sample (n = 211)	Trafficked into sex work (n = 88)	Not trafficked into sex work (n = 123)	
Age, years[a]				.33
≤30 (median)	62.1 (131)	65.9 (58)	59.4 (73)	
>30	37.9 (80)	34.1 (30)	40.7 (50)	
Formal education				.28
No	78.2 (165)	81.8 (72)	75.6 (93)	
Yes	21.8 (46)	18.2 (16)	24.4 (30)	
Marital status				.22
Currently married	9.5 (20)	13.6 (12)	6.5 (8)	
Previously married	49.8 (105)	47.7 (42)	51.2 (63)	
Never married	40.8 (86)	38.6 (34)	42.3 (52)	
Place of origin				.07
Mumbai	6.2 (13)	6.8 (6)	5.7 (7)	
Other parts of Maharashtra	9.0 (19)	13.6 (12)	5.7 (7)	
Karnataka	33.6 (7)	28.4 (25)	37.4 (46)	
Andhra Pradesh	5.2 (11)	5.7 (5)	4.9 (6)	
Bangladesh	6.2 (13)	2.3 (2)	8.9 (11)	
Nepal	20.4 (43)	26.1 (23)	16.3 (20)	
Other	19.4 (41)	17.0 (15)	21.1 (26)	
Religion				.22
Hindu	77.7 (164)	81.8 (72)	74.8 (92)	
Other	22.3 (47)	18.2 (16)	25.2 (31)	
Income per month, rupees[b]				.37
≤3000	63.2 (132)	66.7 (58)	60.7 (74)	
>3000	36.8 (77)	33.3 (29)	39.3 (48)	
Age at entry into sex work				.18
≥18 years	49.8 (105)	44.3 (39)	53.7 (66)	
11–17 years	50.2 (106)	55.7 (49)	46.3 (57)	
Duration in sex work, mean ± SD, years	12.8 ± 5.4	12.7 ± 5.4	12.8 ± 5.4	.86

Abbreviations: HIV, human immunodeficiency virus; SD, standard deviation.

[a] Median age was 30 years.

[b] Median income was 3000 rupees per month.

reports of ≥7 partners per day were classified as high client volume in the first month. Alcohol at first sex work episode was assessed by asking the yes-or-no question "The first time you had sex in exchange for money or gifts, were you drinking alcohol?"

Data Analysis

χ^2 analyses were used to assess whether demographics (all categorical variables) differed based on having been trafficked. Frequencies for being trafficked and for each of the 5 outcome variables (sexual violence, high client volume, nonuse of condoms and frequent alcohol use in first month of sex work, and alcohol use at first sex work episode) were calculated for the total sample. Associations between being trafficked and each outcome variable were assessed via logistic regression, first unadjusted and then adjusted for the following confounders: education, marital status, religion, income, and age at entry into sex work. To minimize the potential for collinearity, we assessed correlation between pairs of independent variables; no variables included in the same regression model were highly correlated (ie, $r > 0.40$). All analyses were performed using SAS software (version 9.1; SAS Institute) [16].

RESULTS

Approximately 2 in 5 (41.7%) HIV-infected FSWs reported being forced or coerced (ie, trafficked) into sex work. The most common perpetrators reported by such women were coworkers (25.0%), acquaintances (25.0%), strangers (18.2%), and family members (9.1%) (data not shown). Slightly more than half (50.2%) entered sex work before age 18 years. The most common places of origin among participating FSWs were

Table 2. Sex Trafficking and Experiences of Violence, Alcohol Use, and HIV Risk in the First Month of Sex Work as Reported by HIV-Infected Female Sex Workers (FSWs) in Mumbai, India (n = 211)

Variable	FSWs, % [95% CI] (No.)
Forced or coerced into sex work	41.7 [35–49] (88)
Sexual violence in first month of sex work	62.2 [56–69] (132)
≥7 clients per day in first month of sex work	35.9 [29–43] (75)
No client condom use during first month of sex work	50.7 [44–58] (107)
Used alcohol very often in first month of sex work	54.0 [47–61] (114)
Used alcohol at first sex work encounter	59.2 [52–66] (125)

Abbreviations: CI, confidence interval; HIV, human immunodeficiency virus.

Mumbai (6.2%), other parts of Maharashtra (9.0%), Karnataka (33.6%), Bangladesh (6.2%), and Nepal (20.4%). Neither age at entry nor other demographics differed significantly between those who reported being trafficked into sex work and those who did not (all $P > .05$) (Table 1). Violence and sexual risk exposures were highly prevalent during the first month in sex work. Sixty-two percent of HIV-infected FSWs reported sexual violence from male clients during the initial month after entry. The volume of male clients during the first month was high, with 35.9% reporting sex with ≥7 clients per day during this period. Slightly more than half (50.7%) of HIV-infected FSWs reported no use of condoms in their first month in sex work. Alcohol related risk was also prevalent; the majority of participants reported frequent alcohol use in the initial month after entry (54.0%) and alcohol use at first sex work episode (59.2%) (Table 2). In multivariate models, those reporting having been trafficked into sex work had significantly higher odds of sexual violence during their initial month in sex work (adjusted odds ratio [AOR] 3.1; 95% confidence interval [CI], 1.6–6.1), a high volume of male clients (AOR, 3.3; 95% CI, 1.8–6.1), and no use of condoms in this same time period (AOR, 3.8; 95% CI, 2.1–7.1). Participating HIV-infected FSWs who reported entering sex work involuntarily also had significantly higher odds of frequent alcohol use during their first month of sex work (AOR, 1.9; 95% CI, 1.0–3.4) and of alcohol use at their first sex work episode (AOR, 2.2; 95% CI, 1.2–4.0) (Table 3).

DISCUSSION

More than 2 in 5 HIV-infected FSWs in the current study reported that they did not enter commercial sex work of their own volition but were instead forced or coerced (ie, trafficked) into the sex trade. This prevalence estimate is slightly greater than what was found in the single other study of sex trafficking among a general sample of FSWs in South Asia (32%) [11]. This greater prevalence is probably due to the high prevalence of HIV infection in the current sample, given that this same earlier study found that being trafficked into sex work was more common among FSWs infected with HIV. Regional differences between Mumbai and West Bengal may also account for this difference. The first month of sex work was characterized by high levels of sexual violence and sexual risk, with almost 2 of 3 FSWs in the current sample reporting such exposure. Those FSWs who were trafficked into sex work had ~3 times the odds of sexual violence at the hands of male clients in the first month after entry, compared with those entering voluntarily. This finding is

Table 3. Associations Between Being Trafficked Into Sex Work and Experiences of Sexual Violence, Frequent Alcohol Use, and HIV Risk in the First Month of Sex Work Among HIV-Infected Female Sex Workers in Mumbai, India (n = 211)

	Female sex workers, % (No.)		OR (95% CI)	
Variable	Trafficked into sex work (n = 88)	Not trafficked into sex work (n = 123)	Crude	Adjusted[a]
Sexual violence in first month of sex work	75.0 (66)	53.7 (66)	2.6 (1.4–4.7)	3.1 (1.6–6.1)
≥7 clients/day in first month of sex work	51.2 (44)	25.2 (31)	3.1 (1.7–5.6)	3.3 (1.8–6.1)
No client condom use in first month of sex work	68.2 (60)	38.2 (47)	3.5 (2.0–6.2)	3.8 (2.1–7.1)
Used alcohol very often in first month of sex work	62.5 (55)	48.0 (59)	1.8 (1.04–1.2)	1.9 (1.0–3.4)
Alcohol use at first sex work episode	69.3 (61)	52.0 (64)	2.1 (1.2–3.7)	2.2 (1.2–4.0)

Abbreviations: CI, confidence interval; HIV, human immunodeficiency virus; OR, odds ratio.

[a] Adjusted ORs were adjusted for history of any formal education, marital status, religion, income, and age at entry into sex work.

consistent with findings of the aforementioned study of FSWs in West Bengal [11] and a similar study of FSWs in Thailand [7], but it provides increased specificity regarding proximity to the time of entry, in that the current assessment of sexual violence was limited to the first 30 days in sex work.

Expanding on previous work assessing links between involuntary entry and HIV risk, current findings indicate that those FSWs who reported being forced or coerced into sex work also experienced far greater early exposure to sexually transmitted infection. Such FSWs had 3 times the odds, compared with those who entered voluntarily, of having had sex with ≥7 partners a day in the first month in sex work, and ~4 times the odds of reporting no condom use by male clients in this same initial period. Alcohol also figured prominently in the early experiences of sex work for trafficked women and girls. Trafficked HIV-infected FSWs had approximately twice the odds of reporting both using alcohol "very often" during the first month in prostitution and using alcohol during their first commercial sex episode. These findings confirm previous qualitative work among trafficked FSWs that describes alcohol being forced on women and girls on entry in order to gain their compliance and voluntary regular use of alcohol subsequent to violent sexual initiation in an attempt to cope with ongoing sexual violence [8]. Relationships between trafficking and other substance use (particularly injection drug use, a major source of HIV infection) warrant further examination.

Most all HIV prevention programs targeting FSWs The high prevalence of trafficking and experiences of sexual violence as part of being forced to participate in sex work described by this sample is critical to recognize; the great majority of current descriptions of conditions of sex work, and HIV prevention programs targeting FSWs assume voluntary entry and autonomy regarding sex work practice [18]. However, the ~40% of women and girls who report being trafficked describe extremely high levels of exposure to HIV risk within the initial month of sex work. Thus, peer education and FSW collectivism (the dominant approaches for HIV prevention among FSWs) are not likely to be effective for a large portion of this population, given that experiences of coercion directly contradict the autonomy and control over sexual protection needed to implement condom use and gain support from fellow FSWs. Further, assuming autonomy may be gradually obtained by those forced or coerced into sex work, the early and intense exposure to infection may well lead them to contract HIV before being able to benefit from such programs. This contention is supported by previous work among trafficked FSWs in Mumbai and in Nepal that found rates of HIV infection between 30% and 60% after only short periods of sex work exposure [9, 15].

There are several limitations related to the design of the current study. Assessments were of experiences in the first month of sex work, leaving open the possibility of recall error. However, recent research indicates that recall for traumatic events is likely to be highly reliable, significantly more than for nontraumatic events [19]. The study surveyed a group of HIV-infected FSWs who were members of an HIV service network; it is possible that the experiences of this population do not reflect those of the larger population of HIV-infected FSWS, or FSWs generally. Finally, the current estimate of the prevalence of trafficking into sex work may be inflated given that being trafficked is associated with elevated risk for HIV infection and the HIV-infected nature of the current sample [11].

In sum, ~2 in 5 HIV-infected FSWs report being trafficked (ie, forced or coerced) into sex work. This experience is associated with increased risk of sexual violence, frequent alcohol use, and high levels of exposure to HIV during the first 30 days after entry. These findings call into question utility of conventional harm reduction approaches to HIV prevention that rely on the autonomy of FSWs. Development of interventions that assure FSW autonomy within the context of sex work should be prioritized. HIV prevention programs and policies should include substantial efforts to prevent the involuntary entry of women and girls into sex work.

Notes

Financial support. This work was supported by the National Institute on Alcohol Abuse and Alcoholism (grants R01AA016059-S1 and K24AA015674) and the Indian Council of Medical Research.

Potential conflicts of interest. All authors: none reported

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. UNAIDS. UNAIDS report on the global AIDS epidemic. Geneva, Switzerland: UNAIDS, **2010**.
2. National AIDS Control Organization (NACO). Annual report: 2008–2009. New Delhi, India: NACO, **2009**.
3. National AIDS Control Organization (NACO). Annual report: 2009–2010. New Delhi, India: NACO, **2010**.
4. World Bank. HIV/AIDS in India. Washington, DC: World Bank, **2007**.
5. National AIDS Control Organization (NACO). HIV sentinel surveillance and HIV estimation in India 2007: a technical brief. New Delhi: NACO, **2007**.
6. US State Department. India: country reports on Human Rights Practices. Washington, DC: US State Department, **2002**.
7. Decker MR, McCauley HL, Phuengsamran D, Janyam S, Silverman JG. Sex trafficking, sexual risk, sexually transmitted infection and reproductive health among female sex workers in Thailand. J Epidemiol Community Health **2011**; Apr;65(4):334–9.
8. Gupta J, Raj A, Decker MR, Reed E, Silverman JG. HIV vulnerabilities of sex-trafficked Indian women and girls. Int J Gynecol Obstet **2009**; 107:30–4.
9. Silverman JG, Decker MR, Gupta J, Maheshwari A, Patel V, Raj A. HIV prevalence and predictors among rescued sex trafficked women and girls in Mumbai, India. J Acquir Immune Defic Syndr **2006**; 43:588–93.
10. Huda S. Sex trafficking in South Asia. Int J Gynaecol Obstet **2006**; 94:274–81.
11. Sarkar K, Bal B, Mukherjee R, et al. Sex-trafficking, violence, negotiating skill, and HIV infection in brothel-based sex workers of eastern India, adjoining Nepal, Bhutan, and Bangladesh. J Health Popul Nutr **2008**; 26:223–31.

12. Samet JH, Pace CA, Cheng DM, et al. Alcohol use and sex risk behaviors among HIV-infected female sex workers and HIV-infected male clients of FSWs in India. AIDS Behav **2010**; 14(Suppl 1):74–83.
13. National AIDS Control Organization (NACO). National Behavioural Surveillance Survey (BSS) 2006: female sex workers (FSWs) and their clients. New Delhi: NACO, **2006**.
14. International Institute for Population Sciences (IIPS). National Family Health Survey (NFHS-3) 2005–2006: India. Mumbai, India: IIPS, **2007**.
15. Silverman JG, Decker MR, Gupta J, et al. Experiences of sex trafficking victims in Mumbai, India. Int J Gynaecol Obstet **2007**; 97:221–6.
16. SAS [computer program]. Version 9.1. Cary, NC: SAS Institute, **2003**.
17. Nair PM. A report on trafficking of women and children in India: 2002–2003. Vol 1. New Delhi, India: UNIFEM, ISS, National Human Rights Commission of India, **2004**.
18. World Health Organization (WHO)/UNAIDS. Addressing violence against women and HIV/AIDS: what works? Geneva, Switzerland: WHO/UNAIDS, **2010**.
19. Lalande KM, Bonanno GA. Retrospective memory bias for the frequency of potentially traumatic events: a prospective study. Psychol Trauma **2011**. In press.

Journal of Gerontology: MEDICAL SCIENCES
Cite journal as: J Gerontol A Biol Sci Med Sci. 2011 July;66A(7):809–814
doi:10.1093/gerona/glr063

Advance Access published on May 31, 2011

Brief Report

Predictors of Thyroid Hormone Initiation in Older Adults: Results From the Cardiovascular Health Study

Lily L. Somwaru,[1] Alice M. Arnold,[2] and Anne R. Cappola[1]

[1]Division of Endocrinology, Diabetes, and Metabolism, University of Pennsylvania School of Medicine, Philadelphia.
[2]Department of Biostatistics, University of Washington, Seattle.

Address correspondence to Anne R. Cappola, MD, ScM, Division of Endocrinology, Diabetes, and Metabolism, University of Pennsylvania School of Medicine, 764CRB, 415 Curie Boulevard, Philadelphia, PA 19104. Email: acappola@mail.med.upenn.edu

Background. Despite widespread use, there are no data on initiation of thyroid hormone use in older people. We report the prevalence of thyroid hormone use and predictors of thyroid hormone initiation in a population of older men and women.

Methods. Thyroid hormone medication data were collected annually from 1989 to 2006 in community-dwelling individuals aged 65 years and older enrolled in the Cardiovascular Health Study (N = 5,888). Associations of age, sex, race, body mass index, education, and coronary heart disease with initiation were evaluated using discrete-time survival analysis.

Results. In 1989–1990, 8.9% (95% confidence interval 8.1%–9.7%) of participants were taking a thyroid hormone preparation, increasing to 20.0% (95% confidence interval 8.2%–21.8%) over 16 years. The average initiation rate was 1% per year. The initiation rate was nonlinear with age, and those aged 85 years and older initiated thyroid hormone more than twice as frequently as those aged 65–69 years (hazard ratio = 2.34; 95% confidence interval 1.43–3.85). White women were more likely to initiate thyroid hormone than any other race and sex group. Higher body mass index was independently associated with higher risk for initiation (p = .002) as was greater education (p = .02) and prevalent coronary heart disease (p = .03).

Conclusions. Thyroid hormone use is common in older people. The indications and benefits of thyroid hormone use in older individuals with the highest rate of thyroid hormone initiation—the oldest old, overweight and obese individuals, and those with coronary heart disease—should be investigated.

Key Words: Thyroid hormone—Levothyroxine—Elderly population.

Received September 2, 2010; Accepted March 11, 2011

Decision Editor: Luigi Ferrucci, MD, PhD

HYPOTHYROIDISM is more common in the elderly population, and the classic signs and symptoms found in younger people are often not apparent (1), making this an attractive demographic for biochemical screening. In addition, screening identifies subclinical hypothyroidism, which is found in up to 15% of older people (2). Management of subclinical hypothyroidism is controversial, and expert panels have published guidelines for (3) and against (4) routine treatment of subclinical hypothyroidism. An increase in the use of routine thyroid-stimulating hormone (TSH) screening and recognition of overt and subclinical hypothyroidism would be anticipated to lead to rising use of thyroid hormone supplementation in the elderly population. To date, there have been no studies to determine if an increase in thyroid hormone use over time has occurred. We conducted an analysis of individuals taking thyroid hormone preparations who were enrolled in the Cardiovascular Health Study (CHS), a population-based study of community-dwelling individuals aged 65 years and older. We sought to describe trends in the prevalence of thyroid hormone use and predictors of thyroid hormone initiation in a population of elderly men and women.

Methods

These analyses are based on data from the CHS (5). The CHS is a population-based, longitudinal study in 5,888 adults 65 years and older. Enrollment of an original cohort of 5,201 adults at four U.S. sites occurred between May 1989 and June 1990, and an additional cohort of 687, predominantly African Americans, was enrolled in 1992–1993. The institutional review boards of all four sites and of the coordinating center at the University of Washington in Seattle approved the study. All participants gave informed consent.

Table 1. Characteristics of Cohort by Thyroid Medication Status at Baseline

	Thyroid Medication			
	Original Cohort		New Cohort	
Characteristic	No (n = 4737)	Yes (n = 464)	No (n = 643)	Yes (n = 44)
Age, mean (*SD*)	72.8 (5.6)	72.9 (5.4)	73.0 (5.8)	73.1 (5.5)
Male, *n* (%)	2,149 (45.4)	90 (19.4)‡	249 (38.7)	7 (15.9)†
Caucasian, *n* (%)	4,476 (94.5)	449 (96.8)*	0	0
High school graduate, *n* (%)	3,393 (71.8)	356 (77.1)*	359 (56.0)	31 (72.1)*
Income ≥$25,000, *n* (%)	1,946 (41.1)	194 (41.8)	137 (21.3)	10 (22.7)
Current smoker, *n* (%)	551 (11.6)	50 (10.8)	95 (14.8)	4 (9.3)
BMI, kg/m^2, mean (*SD*)	26.4 (4.5)	26.7 (4.9)	28.7 (5.6)	29.4 (6.4)
Good or excellent health, *n* (%)	3,643 (77.0)	336 (72.4)*	379 (59.2)	22 (50.0)
Diabetes mellitus, *n* (%)	714 (15.2)	74 (16.1)	153 (25.0)	12 (27.9)
Hypertension, *n* (%)	2,687 (56.8)	255 (55.0)	479 (74.5)	36 (81.8)
Coronary heart disease, *n* (%)	930 (19.6)	92 (19.8)	118 (18.4)	14 (31.8)*
Stroke, *n* (%)	181 (3.8)	18 (3.9)	47 (7.3)	3 (6.8)
Any ADL difficulty, *n* (%)	332 (7.0)	42 (9.1)	93 (14.5)	9 (20.5)

Notes: ADL = activities of daily living; BMI = body mass index.
*$p < .05$; †$p < .01$; ‡$p < .001$ for comparison by medication use within cohort.

Thyroid hormone medication use was assessed annually from Study Year 1 (1989–1990) through Study Year 10 (1998–1999) via medication bottle examination during annual study visits and from Study Years 11 to 17 (1999–2000 to 2005–2006) by annual surveillance phone calls. Body mass index (BMI) was calculated as kg/m^2 using objective measures. Diabetes mellitus was classified using the American Diabetes Association criteria (6). Coronary heart disease (CHD) was present if one of the following was reported and confirmed by adjudication: myocardial infarction, angina pectoris, or history of angioplasty or bypass surgery (7). Stroke diagnosis was also confirmed by adjudication.

Statistical Analysis

Study participants' baseline characteristics were summarized by study cohort, and participants taking a thyroid hormone preparation at baseline were compared with those who were not using a *t*-test or chi-square test as appropriate. The percentage taking thyroid hormone medication was calculated by year, overall and stratified by sex and race. Annual initiation rates were calculated after exclusion of 508 participants taking thyroid hormone at baseline. Time to initiation was defined as the number of years after baseline that thyroid medication use was first reported. Discrete-time survival analysis was used to evaluate associations of baseline age, sex, race, education, income, smoking, BMI, weight gain, self-reported health, fatigue, diabetes, hypertension, prevalent CHD, stroke, and difficulty in activities of daily living with initiation of use. Only statistically significant ($p < .05$) variables were retained in the final model. Age and BMI were modeled continuously, and results shown on both continuous and categorical scales, with p values for all variables derived from the continuous model. Participants were censored at the time of their last visit. All analyses were performed using STATA version 9 (Stata Corp., College Station, TX).

Results

The mean age was 72.8 years (range 65–100 years), 58% were women and 84% white. At baseline, thyroid hormone users were more likely to be women, Caucasian, and high school graduates than nonusers were (Table 1). They were less likely to self-report good or excellent health, which was statistically significant in the original cohort, and more likely to have CHD only in the minority cohort.

In 1989–1990, 8.9% (95% confidence interval [CI] 8.1%–9.7%) of participants were taking a thyroid hormone preparation, increasing to 20.0% (95% CI 18.2%–21.8%) over 16 years, by 2005–2006. More thyroid hormone use was seen in women than in men, with a greater proportion of users at each year between 1989 and 1990 and 2005 and 2006 (Figure 1), and a greater proportion of users in whites than in nonwhites. At initiation of the study, 12.9% (95% CI 11.7%–14.2%) of white women were taking thyroid hormone, increasing to 26.3% (95% CI 23.5%–29.0%) by 2005–2006. In nonwhite women, 7.6% (95% CI 3.6%–11.5%) were taking thyroid hormone at enrollment, increasing to 13.2% (95% CI 8.9%–17.5%) by 2005–2006. Despite the lower proportion of thyroid hormone use across all years, white and nonwhite men also demonstrated a trend of increasing thyroid hormone use over calendar time. Only 4.1% (95% CI 3.3%–5.0%) of white men and 1.9% (95% CI 0.01%–4.1%) of nonwhite men were taking a thyroid hormone preparation at baseline, increasing to 13.6% (95% CI 10.7%–16.6%) and 8.2% (95% CI 2.7%–13.6%), respectively, by the end of follow-up.

After excluding the 508 thyroid medication users at baseline, there was a rate of initiation of 0.6%–1.4% per year among nonusers, with a total of 498 people initiating thyroid hormone after baseline. The average rate of initiation was 1% per year and without a consistent linear trend with time. The average age at thyroid hormone initiation was 79.7 ± 6.4 years. However, the association with age

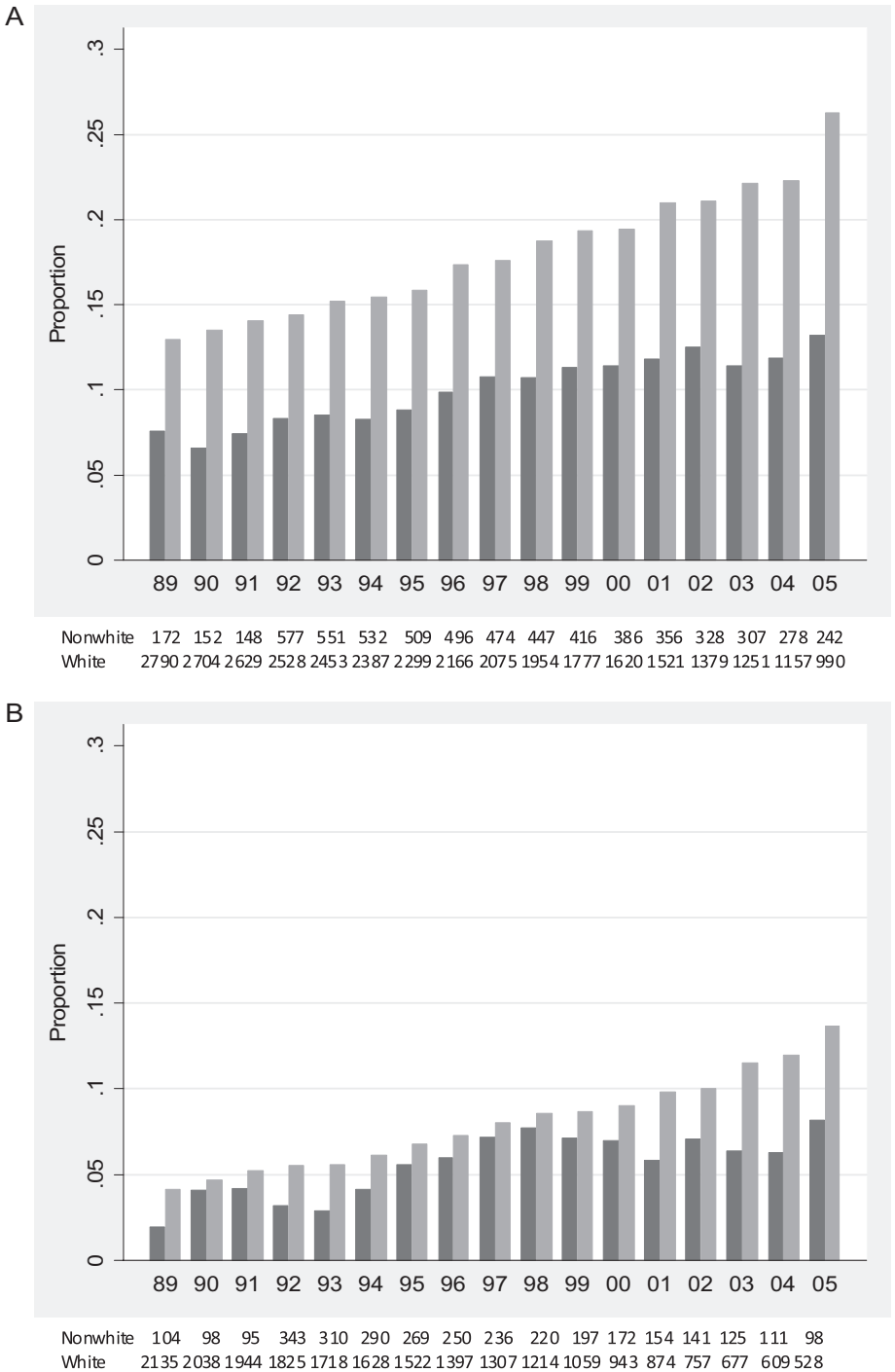


Figure 1. Proportion of participants taking thyroid hormone medication by calendar year: (A) white and nonwhite women and (B) white and nonwhite men. Black bars indicate nonwhites, gray bars whites.

was not linear, with a higher crude incidence rate, at 1.50% per year, in those aged 85 years and older at baseline than in the younger age groups, in which the crude incidence rate varied from 0.80% to 0.88% per year. In multivariable analyses (Table 2), those aged 85 years and older at baseline were more than twice as likely to initiate thyroid hormone (hazard ratio 2.34; 95% CI 1.43–3.85) than those aged 65–69 years. White women were more likely to initiate thyroid hormone than any other race and sex group (overall p value <.001), with no significant difference in rates of thyroid hormone initiation among white men, nonwhite women, and nonwhite men. Thyroid hormone initiation was higher in those whose BMI was above 25 kg/m^2 than below, achieving statistical significance for

Table 2. Hazard Ratios (95% confidence intervals [CI]) From Discrete-Time Survival Analysis of Initiation of Thyroid Medication Use

Characteristic	*n*	Hazard Ratio	95% CI	*p* Value*
Age at baseline, y	—	1.02	1.00–1.04	.019
65—69	1,825	1.00	Reference	—
70—74	1,664	1.05	0.85–1.30	—
75—79	1,051	1.05	0.81–1.36	—
80—84	493	1.26	0.88–1.82	—
85+	188	2.34	1.43–3.85	—
Sex/race group	—	—	—	<.001
White women	2,385	1.00	Reference	—
White men	1,972	0.61	0.50–0.75	—
Nonwhite women	535	0.42	0.28–0.63	—
Nonwhite men	329	0.52	0.32–0.85	—
BMI, kg/m^2	—	1.03	1.01–1.05	.002
Normal (<25.0)	2,005	1.00	Reference	—
Overweight (25.0–29.9)	2,170	1.26	1.03–1.54	—
Obese (30.0–39.9)	959	1.25	0.96–1.62	—
Extremely obese (>40.0)	68	2.72	1.45–5.10	—
Education, y	—	1.03	1.00, 1.07	.022
CHD at baseline	988	1.31	1.04, 1.65	.028

Notes: BMI = body mass index; CHD = coronary heart disease.

*The *p* values shown are derived from a model in which each variable was modeled continuously. Results for age and BMI using categorical scales are also displayed.

the overweight (hazard ratio 1.26; 95% CI 1.03–1.54) and extremely obese (hazard ratio 2.72; 95% CI 1.45–5.10) groups. Initiation of thyroid hormone was also independently associated with more years of education ($p = .022$) and CHD at baseline ($p = .028$).

Conclusions

We found a high proportion of thyroid hormone users in our cohort of community-dwelling individuals aged 65 years and older, particularly among white women. Our study is the first to follow patients over an extended time frame, demonstrating a steady trend in thyroid hormone initiation. We also provide new demographic information about who initiates thyroid hormone, with age 85 years or older, being a white woman, more years of education, high BMI, and prevalent CHD independently predicting thyroid hormone initiation over our 16-year time frame of study.

Our study demonstrates a similar prevalence of thyroid hormone use by sex and race to that reported in the 70- to 79-year-olds enrolled in the Health, Aging, and Body Composition Study in 1997–1999 and to a white, community-dwelling population of women aged 65 years or older enrolled in the Study of Osteoporotic Fractures in 1986–1988, during each of these time frames (8,9). Hashimoto's thyroiditis is more common in women than in men, and thus, our finding of greater thyroid hormone use in women is not surprising and parallels this sex difference in indication for prescription of thyroid hormone. The differences we found in thyroid hormone use by sex and race correspond to reported findings of demographic differences in TSH distribution (10–13) and suggest that bias in screening practices by sex and race likely play a minor role. We also found a higher thyroid hormone initiation rate with increasing educational level and with CHD, which could suggest higher screening rates in more educated individuals and in those with cardiovascular disease.

We found that a BMI above normal is also associated with increased initiation of thyroid hormone preparations in this elderly cohort. A higher prevalence of subclinical hypothyroidism has been shown in obesity (14), and it is likely that individuals with concerns about their weight were more likely to have thyroid function testing performed and, in turn, to be prescribed thyroid hormone replacement. Interestingly, weight loss after bariatric surgery has been shown to reverse subclinical hypothyroidism in obese younger individuals (15). These data suggest that obesity depletes thyroid reserves, resulting in subclinical hypothyroidism, rather than the converse effect of mild thyroid dysfunction inducing weight gain.

Our most intriguing finding is a higher rate of thyroid hormone initiation among those aged 85 years or older that is independent of sex or race. Although it is possible that there is an increase in overt hypothyroidism in this age-group, the more likely explanation is prescription of thyroid hormone for treatment of subclinical hypothyroidism, which is present in 14.5% of the population of men and women aged 80 years and older (10). However, the benefits of thyroid hormone supplementation of subclinical hypothyroidism are unclear in this age-group. Data from the Leiden 85+ study show lower mortality in 85-year-old men and women with elevated TSH levels compared with their euthyroid counterparts and no difference in functional status (16). Offspring of nonagenarians tend to have higher TSH levels than their partners do, also suggesting a favorable effect of slower thyroid metabolism on longevity (17). A small study of nonagenarians showed no association between TSH level and mortality, although only 4% had elevated TSH levels (18). Furthermore, we have previously found in CHS that overreplacement with thyroid hormone is common in older people (19), and in Study of Osteoporotic Fractures, thyroid hormone use was independently associated with greater declines in lower extremity performance (20). Furthermore, we and others have previously found that overreplacement with thyroid hormone is common in older people (21,22), and others have shown adverse cardiac, skeletal, and cognitive effects in older people with low TSH levels (23–25). In aggregate, these data suggest the need to define a target TSH for thyroid hormone initiation, risk–benefit ratio of thyroid hormone replacement, and therapeutic goals that are specific to the 85 years and older age-group. This need is only highlighted by the fact that those aged 85 years and older are the fastest growing demographic group (26), and

a group in whom polypharmacy is a serious problem (27,28).

A major strength of this study is the use of a large, population-based cohort of older men and women to examine trends in thyroid hormone replacement over 16 years. Thyroid function testing was not performed in the CHS main study, and when it was performed using banked samples, results were never released to participants or their physicians. Thus, participation in CHS should not have influenced the prescribing patterns of participant's physicians. Limitations of our study include the lack of information on thyroid function testing prior to thyroid hormone initiation or the prescriber's indication for thyroid hormone prescription, and use of baseline covariates in the models. We were also unable to provide data after the 2005–2006 participant phone call.

Implications

Levothyroxine sodium was the fourth most commonly dispensed medication in the United States in 2008 (29). Mild TSH elevations increase in prevalence with increasing age, particularly in those aged 70 years or older (30). The management of subclinical hypothyroidism in the elderly population is controversial, with observational studies largely showing no harm in those with TSH levels lower than 10 mU/L (31,32) and no data from randomized clinical trials with clinical outcomes. Our data support the need to further investigate the threshold TSH level for thyroid hormone initiation and benefits of thyroid hormone use in the elderly population, particularly in the oldest old (aged 85 years and older), overweight and obese individuals, and those with CHD, who have the highest rates of thyroid hormone initiation.

Funding

This work was supported by the National Institute on Aging at the National Institutes of Health (R01-AG032317). The Cardiovascular Health Study was supported by contract numbers N01-HC-80007, N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133 and grant number U01 HL-080295 from the National Heart, Lung, and Blood Institute, with additional contribution from the National Institute of Neurological Disorders and Stroke. Additional support was provided through R01-AG-015928, R01-AG-020098, R01-AG-023629, and R01-AG-027058 from the National Institute on Aging; R01 HL-075366 from the National Heart, Lung, and Blood Institute; and P30-AG-024827 from the University of Pittsburgh Claude. D. Pepper Older Americans Independence Center. A full list of principal CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm.

References

1. Doucet J, Trivalle C, Chassagne P, et al. Does age play a role in clinical presentation of hypothyroidism? *J Am Geriatr Soc*. 1994;42:984–986.
2. Cappola AR, Fried LP, Arnold AM, et al. Thyroid status, cardiovascular risk, and mortality in older adults. *JAMA*. 2006;295:1033–1041.
3. Gharib H, Tuttle RM, Baskin HJ, Fish LH, Singer PA, McDermott MT. Subclinical thyroid dysfunction: a joint statement on management from the American Association of Clinical Endocrinologists, the American Thyroid Association, and the Endocrine Society. *J Clin Endocrinol Metab*. 2005;90:581–585.
4. Surks MI, Ortiz E, Daniels GH, et al. Subclinical thyroid disease: scientific review and guidelines for diagnosis and management. *JAMA*. 2004;291:228–238.
5. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991;1:263–276.
6. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*. 1997;20:1183–1197.
7. Psaty BM, Kuller LH, Bild D, et al. Methods of assessing prevalent cardiovascular disease in the Cardiovascular Health Study. *Ann Epidemiol*. 1995;5:270–277.
8. Kanaya AM, Harris F, Volpato S, Perez-Stable EJ, Harris T, Bauer DC. Association between thyroid dysfunction and total cholesterol level in an older biracial population: the health, aging and body composition study. *Arch Intern Med*. 2002;162:773–779.
9. Bauer DC, Rodondi N, Stone KL, Hillier TA. Thyroid hormone use, hyperthyroidism and mortality in older women. *Am J Med*. 2007;120:343–349.
10. Hollowell JG, Staehling NW, Flanders WD, et al. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab*. 2002;87:489–499.
11. Boucai L, Surks MI. Reference limits of serum TSH and free T4 are significantly influenced by race and age in an urban outpatient medical practice. *Clin Endocrinol (Oxf)*. 2009;70:788–793.
12. Hamilton TE, Davis S, Onstad L, Kopecky KJ. Thyrotropin levels in a population with no clinical, autoantibody, or ultrasonographic evidence of thyroid disease: implications for the diagnosis of subclinical hypothyroidism. *J Clin Endocrinol Metab*. 2008;93:1224–1230.
13. Bagchi N, Brown TR, Parish RF. Thyroid dysfunction in adults over age 55 years. A study in an urban US community. *Arch Intern Med*. 1990;150:785–787.
14. Asvold BO, Bjoro T, Vatten LJ. Association of serum TSH with high body mass differs between smokers and never-smokers. *J Clin Endocrinol Metab*. 2009;94:5023–5027.
15. Chikunguwo S, Brethauer S, Nirujogi V, et al. Influence of obesity and surgical weight loss on thyroid hormone levels. *Surg Obes Relat Dis*. 2007;3:631–635.
16. Gussekloo J, van EE, de Craen AJ, Meinders AE, Frolich M, Westendorp RG. Thyroid status, disability and cognitive function, and survival in old age. *JAMA*. 2004;292:2591–2599.
17. Rozing MP, Westendorp RG, de Craen AJ, et al. Low serum free triiodothyronine levels mark familial longevity: the Leiden Longevity Study. *J Gerontol A Biol Sci Med Sci*. 2010;65:365–368.
18. Formiga F, Ferrer A. Thyrotropin serum values and 3-year mortality in nonagenarians. *J Gerontol A Biol Sci Med Sci*. 2010;65:1250–1251.
19. Simonsick EM, Newman AB, Ferrucci L, et al. Subclinical hypothyroidism and functional mobility in older adults. *Arch Intern Med*. 2009;169:2011–2017.
20. Forrest KY, Zmuda JM, Cauley JA. Correlates of decline in lower extremity performance in older women: a 10-year follow-up study. *J Gerontol A Biol Sci Med Sci*. 2006;61:1194–1200.
21. Somwaru LL, Arnold AM, Joshi N, Fried LP, Cappola AR. High frequency of and factors associated with thyroid hormone over-replacement and under-replacement in men and women aged 65 and over. *J Clin Endocrinol Metab*. 2009;94:1342–1345.
22. Diez JJ. Hypothyroidism in patients older than 55 years: an analysis of the etiology and assessment of the effectiveness of therapy. *J Gerontol A Biol Sci Med Sci*. 2002;57:M315–M320.
23. Sawin CT, Geller A, Wolf PA, et al. Low serum thyrotropin concentrations as a risk factor for atrial fibrillation in older persons. *N Engl J Med*. 1994;331:1249–1252.

24. Bauer DC, Ettinger B, Nevitt MC, Stone KL. Risk for fracture in women with low serum levels of thyroid-stimulating hormone. *Ann Intern Med*. 2001;134:561–568.
25. Ceresini G, Lauretani F, Maggio M, et al. Thyroid function abnormalities and cognitive impairment in elderly people: results of the Invecchiare in Chianti study. *J Am Geriatr Soc*. 2009;57:89–93.
26. U.S. Census Bureau. *U.S. Interim Projections by Age, Sex, Race, and Hispanic Origin*. 2004. U.S. Census Bureau Web site. http://www.census.gov/ipc/www/usinterimproj. Accessed March 18, 2004.
27. Espino DV, Bazaldua OV, Palmer RF, et al. Suboptimal medication use and mortality in an older adult community-based cohort: results from the Hispanic EPESE Study. *J Gerontol A Biol Sci Med Sci*. 2006;61:170–175.
28. Bernabei R, Caputi A, Di CL, et al. Need for redesigning pharmacologic research in older individuals. A position statement of the Geriatric Working Group of the Agenzia Italiana del Farmaco (AIFA). *J Gerontol A Biol Sci Med Sci*. 2011;66:66–67.
29. IMS Health. *IMS National Sales Perspectives*. 2009. www.imshealth.com. Accessed August 10, 2010.
30. Surks MI, Hollowell JG. Age-specific distribution of serum thyrotropin and antithyroid antibodies in the US population: implications for the prevalence of subclinical hypothyroidism. *J Clin Endocrinol Metab*. 2007;92:4575–4582.
31. Rodondi N, den Elzen WP, Bauer DC, et al. Subclinical hypothyroidism and the risk of coronary heart disease and mortality. *JAMA*. 2010;304:1365–1374.
32. Rodondi N, Bauer DC, Cappola AR, et al. Subclinical thyroid dysfunction, cardiac function, and the risk of heart failure. The Cardiovascular Health Study. *J Am Coll Cardiol*. 2008;52:1152–1159.

JOURNAL OF VIRAL HEPATITIS
JVH

Journal of Viral Hepatitis, 2011, **18**, 474–481 doi:10.1111/j.1365-2893.2010.01327.x

Hepatitis C testing practices and prevalence in a high-risk urban ambulatory care setting

W. N. Southern,[1,2,3] M.-L. Drainoni,[4,5] B. D. Smith,[6] C. L. Christiansen,[4] D. McKee,[7] A. L. Gifford,[4,8,9] C. M. Weinbaum,[6] D. Thompson,[10] E. Koppelman,[4,5] S. Maher[7] and A. H. Litwin[1,2]

[1]*Department of Medicine, NY, USA;* [2]*Division of General Internal Medicine, NY, USA;* [3]*Section of Hospital Medicine, NY, USA;* [4]*Health and Policy Management Department, School of Public Health, Boston University, NY, USA;* [5]*Center for Health Quality, Outcomes and Economic Research, ENRM Veterans Administration Hospital, NY, USA;* [6]*Division of Viral Hepatitis, Centers for Disease Control and Prevention, National Center for HIV/Viral Hepatitis/STD/TB Prevention, GA, USA;* [7]*Department of Family Medicine; Albert Einstein College of Medicine, Montefiore Medical Center, NY, USA;* [8]*Section of General Internal Medicine, Boston University School of Medicine, NY, USA;* [9]*VA QUERI-HIV/Hepatitis Program, Edith Nourse Rogers Memorial Veterans Hospital, NY, USA; and* [10]*Emerging Health Information Technology, NY, USA*

Received January 2010; accepted for publication March 2010

SUMMARY. Approximately 3.2 million persons are chronically infected with the hepatitis C virus (HCV) in the U.S.; most are not aware of their infection. Our objectives were to examine HCV testing practices to determine which patient characteristics are associated with HCV testing and positivity, and to estimate the prevalence of HCV infection in a high-risk urban population. The study subjects were all patients included in the baseline phase of the Hepatitis C Assessment and Testing Project (HepCAT), a serial cross-sectional study of HCV screening strategies. We examined all patients with a clinic visit to Montefiore Medical Center from 1/1/08 to 2/29/08. Demographic information, laboratory data and ICD-9 diagnostic codes from 3/1/97–2/29/08 were extracted from the electronic medical record. Risk factors for HCV were defined based on birth date, ICD-9 codes and laboratory data. The prevalence of HCV infection was estimated assuming that untested subjects would test positive at the same rate as tested subjects, based on risk-factors. Of 9579 subjects examined, 3803 (39.7%) had been tested for HCV and 438 (11.5%) were positive. The overall prevalence of HCV infection was estimated to be 7.7%. Risk factors associated with being tested and anti-HCV positivity included: born in the high-prevalence birth-cohort (1945–64), substance abuse, HIV infection, alcohol abuse, diagnosis of cirrhosis, end-stage renal disease, and alanine transaminase elevation. In a high-risk urban population, a significant proportion of patients were tested for HCV and the prevalence of HCV infection was high. Physicians appear to use a risk-based screening strategy to identify HCV infection.

Keywords: hepatitis C, prevalence, screening.

INTRODUCTION

An estimated 3.2 million persons are chronically infected with the hepatitis C virus (HCV) in the U.S. [1], roughly three times as many as are infected with HIV [2]. HCV infection is thought to cause approximately 40% of chronic liver disease [3] and the majority of hepato-cellular carcinoma [4] Although the prevalence of anti-HCV is estimated at 1.6% in the U.S. [1], urban populations bear a disproportionate burden of infection and inner city prevalence has been reported as high as 8.3% [5]. Effective treatment for HCV infection is available [6–10], but the majority of those infected are not aware of their status [11–15]. Although testing for patients at high risk is recommended [3,9,10,16,17], optimal testing strategies have not been described [18]. To inform the discussion of testing strategies, we sought to examine the associations between patient characteristics and HCV testing practices among physicians, and estimate the prevalence of HCV infection in a high-risk urban population.

It has been suggested that routine testing for HCV is not efficient [17] or cost-effective [19,20]. Guidelines suggest testing patients with a history of transfusion or organ transplant prior to 1992, persons using injection drugs [3,9,16,17], those with HIV infection [3,9,10], those receiving hemodialysis [3,9,16,17], children of HCV-infected

Abbreviations: ALT, alanine transaminase; EMR, electronic medical record; HCV, hepatitis C virus; MMC, Montefiore Medical Center.

Correspondence: William N. Southern, MD, Department of Medicine, Montefiore Medical Center, 111 East 210th Street, Bronx, NY 10467, USA. E-mail: wsouther@montefiore.org

mothers, and persons with unexplained elevated alanine transaminase (ALT) levels [3,9,17]. In addition, it has been noted that prevalence of HCV infection is very high in patients with a history of alcohol abuse [21,22], sexually-transmitted diseases (STD) [23–25], and psychiatric disease [26–29]. It has also been noted that the majority of prevalent cases of HCV infection are found in patients born between 1945–1964 [1,30,31], and thus, being born in this high prevalence birth-cohort may be considered a risk factor for HCV infection.

It is unclear which of these potential risk factors physicians consider important when deciding which patients to test for HCV, and which testing strategies yield high rates of positivity. The objectives of this analysis were to examine the testing practices of physicians to determine which patient characteristics are associated with testing for HCV antibody and HCV infection, and to estimate the prevalence of HCV infection in a high-risk urban population. We hypothesized that many patient risk factors would be independently associated with HCV testing, and that the prevalence of HCV infection in this population would be significantly higher than the national prevalence.

METHODS

Study setting

The study was conducted at three community-based primary care (family medicine or internal medicine) clinics affiliated with Montefiore Medical Center (MMC), a university-affiliated teaching hospital. The three participating primary care clinics are large, urban clinics located in the Bronx, New York. Each year, 54 000 adults make over 150 000 primary care visits to the three clinics. The clinic sites are located in economically depressed areas of the Bronx and serve patients with high rates of poverty and substance use. Reported prevalence of HCV infection is higher in New York City [32] than the national estimate and the Bronx has a higher prevalence than NYC as a whole [33].

Study design

This study employed a cross-sectional design with retrospective electronic medical record (EMR) review to examine the associations between patient demographic and clinical characteristics, testing for anti-HCV, and anti-HCV positivity.

Study population

All study subjects were patients included in the baseline testing phase of the Hepatitis C Assessment and Testing Project (HepCAT), a serial cross-sectional intervention study investigating the optimal strategy to improve screening for HCV. A qualifying visit was defined as a primary care visit by patients 18 years and older to one of the three participating clinics between 1/1/08 to 2/29/08.

Data extraction

For research and quality improvement purposes, MMC maintains a data replicate of its computerized Clinical Information System containing patient demographics, outpatient visit records, hospital records, ICD-9 codes, prescriptions, and laboratory test results. From this replicate, we extracted demographic information associated with the qualifying clinic visit for each subject. In addition, we extracted clinical information dating back to March 1997, the year electronic records became available, including inpatient and outpatient ICD-9 diagnosis codes, prescription and inpatient medication records, and laboratory testing results. The Institutional Review Boards of Boston University Medical Center and MMC approved this study. Because the dataset contains only de-identified records, informed consent was not obtained from patients or physicians; instead, a Health Insurance Portability and Acountability Act-approved data use agreement [34,35] was signed by all participating investigators.

Outcome variables

For the current analysis, the primary outcomes were ''ever tested'' for HCV antibody and HCV antibody positivity. Ever tested for HCV was defined as an anti-hepatitis C virus antibody (anti-HCV) by ELISA performed from March 1997 through May 2008. HCV antibody positivity (indicating past or current HCV infection) was defined as a positive anti-HCV test from March 1997 through May 2008.

Independent variables/definitions

The major independent variables were demographic and clinical patient characteristics shown to be associated with HCV antibody positivity. Although a history of blood transfusion or organ transplant before 1992 is a known risk factor for HCV infection, the EMR had little data on these risks, so the analysis does not include them. In order to create clinically meaningful diagnosis groups, ICD-9 codes were classified using the Healthcare Cost and Utilization Project of the Agency for Healthcare Research and Quality system [36].

Age

For analysis, age was categorized into five distinct groups. In addition, age was dichotomized as within the high prevalence birth cohort (born from 1945 through 1964) defined by the Centers for Disease Control and Prevention (CDC) [1,30] *vs* not within the cohort.

Sex
Dichotomized as male and female.

Race/Ethnicity
For analysis, race/ethnicity was collapsed into four categories: non-Hispanic White, non-Hispanic Black or African American, Latino or Hispanic, and other/unknown.

Substance abuse
Substance abuse was coded as present if an ICD-9 code for substance abuse/dependence or a positive urine toxicology for amphetamines, barbiturates, cocaine, or methadone was recorded at any time from March 1997 through the qualifying visit date.

HIV
HIV was coded as present if an ICD-9 code for HIV infection or a positive antibody test confirmed by a Western blot was present at any time from March 1997 through the qualifying visit date.

Sexually transmitted disease
Sexually transmitted disease was coded as present if an ICD-9 code indicating gonorrhea or chlamydia or positive gonorrhea or chlamydia PCR probe was present at any time from March 1997 through the qualifying visit date.

Alcohol abuse
Alcohol abuse was coded as present if an ICD-9 code for alcohol dependence or alcohol-related liver disease, or a serum alcohol level ≥ 80 mg/dL was present at any time from March 1997 through the qualifying visit date.

Cirrhosis
Cirrhosis was coded as present if an ICD-9 code for cirrhosis was present at any time from March 1997 through the qualifying visit date.

End stage renal disease
Coded as present if an ICD-9 code for end-stage renal disease or procedure code for hemodialysis was present at any time from March 1997 through the clinic visit date.

Psychiatric disease
Coded as present if an ICD-9 code for affective disorder, anxiety disorder, schizophrenia, or psychosis was present at any time from March 1997 through the clinic visit date.

Alanine transaminase elevation
The highest ALT value reported from March 1997 through the clinic visit date for each subject was used. ALT was treated as a dichotomous variable: >40 U/L was defined as elevated (40 U/L is a commonly used upper limit of normal [37,38]).

Statistical analysis

Estimating the prevalence of hepatitis C virus infection
Although not all subjects were tested for HCV we estimated floor and ceiling values for the prevalence of HCV infection in our population. The floor estimation assumed that all untested subjects were negative. The ceiling estimation was calculated as follows: a predictive logistic regression model was constructed using the tested population to assign a probability of positivity based on co-morbidities associated with positivity. Assuming that untested subjects would test positive at the same rate as tested subjects based on risk profile, this predictive model was applied to the untested population to assign a probability of positivity in each untested subject. The sum of the untested subjects' probabilities was used to estimate the number of subjects who would have tested positive in the untested population.

Proportion tested/proportion positive
The proportion of patients tested for anti-HCV and the proportion of patients testing positive are reported. The proportions tested and positive were calculated for predefined age categories and demographic characteristics, presence or absence of pre-defined co-morbidities, and the presence or absence of ALT elevation.

To examine the relationship between subject age, and other demographic characteristics, co-morbidities and ALT levels, we calculated the proportion of subjects testing positive in each age category stratified by demographics, co-morbidities, and ALT categories.

To examine factors independently associated with HCV testing, a multivariate logistic regression model was constructed; factors eligible for the model included demographics (age, sex, race/ethnicity), high-risk co-morbidities (substance abuse, alcohol abuse, HIV, STD, cirrhosis, end-stage renal disease, psychiatric disease), and ALT elevation. The model was constructed in a forward stepwise fashion including each factor that maintained an independent association with anti-HCV testing (Wald statistic $P < 0.10$). A similar logistic regression model was constructed to examine factors independently associated with testing positive for anti-HCV.

STATA/IC software, version 10.0, (StataCorp, College Station, TX, USA) was used for all data management and statistical analysis.

RESULTS

Study population

Data on 9579 patients were examined. Demographic and clinical information for the study population are summarized in Table 1. The mean age was 48.6 years (range 18–101). The study population was predominantly female (72.4%) and predominantly Latino (51.3%) or African American

Table 1 Characteristics of study population

	(*n* = 9579)
Age	48.6 ± 16.9
Male	2647 (27.6)
Race/Ethnicity	
White	471 (4.9)
Black	3038 (31.7)
Latino	4915 (51.3)
Oth/Unknown	1155 (12.1)
Diagnoses	
Substance Abuse*	558 (5.8)
Alcohol Abuse†	171 (1.8)
HIV‡	429 (4.5)
STD§	271 (2.8)
Cirrhosis¶	97 (1.0)
ESRD**	74 (0.8)
Psychiatric diagnosis††	1550 (16.2)

Continuous variables reported as mean ± standard deviation dichotomous variables reported as No. (%). *ICD-9 or positive urine toxicology. †ICD-9 for Etoh dependence or etoh liver disease or etoh level ≥ 80. ‡ICD-9 or positive antibody test or western blot. §STD, Sexually Transmitted Disease (not HIV): ICD-9 or Positive GC or Chlamydia PCR probe. ¶ICD-9 Code. **ESRD, End-Stage Renal Disease: ICD-9 code or procedure code for hemodialysis. ††ICD-9 for affective, anxiety, schizophrenia, or psychosis.

(31.7%). History of psychiatric disease was reported for 1550 (16.2%) subjects, 558 (5.8%) had a history of substance abuse, and 429 (4.5%) had a history of HIV.

Estimated prevalence

Anti-HCV prevalence among the 3803 (39.7%) persons in this sample tested in our medical systems was 11.5%. The floor estimate of HCV prevalence for the entire study population (assuming all untested subjects are negative) was 4.6%. The ceiling estimate of HCV prevalence (assuming untested subjects would test positive at the same rate as those tested, based on risk profile) was 7.7%.

Hepatitis C testing by age, high risk diagnosis, and alanine transaminase elevation

The proportion of patients tested for anti-HCV and the proportion testing positive stratified by demographics, high-risk co-morbidities, and ALT elevation are reported in Table 2. Several high risk co-morbidities were associated with a large proportion of subjects tested including substance abuse (78.1% tested, 43.8% positive), alcohol abuse (74.3% tested, 33.1% positive), HIV (87.4% tested, 34.4% positive), cirrhosis (89.7% tested, 51.7% positive), and end-stage renal disease (85.1% tested, 9.5% positive). A substantial proportion of subjects aged 18–29 years were tested (30.3%), but a small proportion of those tested positive (0.4%). Of subjects with any risk factor (in the high-prevalence birth cohort, any high-risk co-morbidity, or elevation of ALT), 48.6% were tested and 15.7% of those tested positive. Of subjects without any risk factor noted, 28.8% were tested, and of those, 3.0% were positive.

Multivariate analysis of testing

Bivariate and multivariate associations between factors and HCV testing are reported in Table 3. In multivariate analysis, each of the following factors was significantly independently associated with anti-HCV testing: born in high prevalence birth cohort; male sex; African-American race; Latino ethnicity; substance abuse; alcohol abuse; HIV; STD; cirrhosis; end-stage renal disease; psychiatric disease; and elevation of ALT.

Multivariate analysis of testing positive

Bivariate and multivariate associations between factors and testing positive for anti-HCV are reported in Table 4. In multivariate analysis each of the following factors was significantly independently associated with testing positive for anti-HCV: born in high prevalence birth cohort; male sex; substance abuse; HIV; cirrhosis; and elevation of ALT.

DISCUSSION

Testing practices in the three clinics evaluated in this study show that physicians test patients with known risk factors to identify HCV infection. The majority of patients with substance abuse (78.1%), alcohol abuse (74.3%), HIV (87.4%), cirrhosis (89.7%), end-stage renal disease (85.1%), ALT elevation (67.2%), or STDs (52.8%) were tested. In addition, a substantial proportion of patients with psychiatric diagnosis (49.7%) were tested. Each of these factors was independently associated with testing in multivariate analysis.

The majority of anti-HCV positive patients identified (73.3%) were born in the high prevalence birth-cohort. Being born in these years was also independently associated with HCV testing and anti-HCV positivity in multivariate analysis. Although testing all patients born in the high prevalence birth cohort may be warranted, evidence suggests that birth cohort-based testing alone would be a less than optimal strategy. First, our data suggest that birth cohort-based testing would fail to identify 26.7% of anti-HCV positive persons, which is similar to the unidentified proportions found when testing only in the birth cohort reported by O'Brien (25.4%) [31], Armstrong (34.4%) [1], and Alter (31.3%) [30]. Second, several factors were independently and strongly associated with positivity after

Table 2 Hepatitis C testing stratified by demographic characteristics, co-morbidities, and ALT elevation (n = 9579)

	Tested No. (%)	Positive No. (%)
Demographics		
Age		
18–29 (n = 1571)	476 (30.3)	2 (0.4)
30–44 (n = 2443)	1006 (41.2)	61 (6.1)
45–54 (n = 2050)	999 (48.7)	173 (17.3)
55–64 (n = 1644)	737 (44.8)	148 (20.1)
≥65(n = 1871)	585 (31.3)	54 (9.2)
Sex		
Male (n = 2647)	1297 (49.0)	239 (18.4)
Female (n = 6932)	2506 (36.2)	199 (7.9)
Race/Ethnicity		
White (n = 471)	198 (42.0)	36 (18.2)
African American (n = 3038)	1244 (40.9)	133 (10.7)
Latino (n = 4915)	1966 (40.0)	242 (12.3)
Oth/Unknown (n = 1155)	395 (34.2)	27 (6.8)
High-risk co-morbidities		
Substance Abuse* (n = 558)	436 (78.1)	191 (43.8)
Etoh Abuse† (n = 171)	127 (74.3)	42 (33.1)
HIV‡ (n = 429)	375 (87.4)	129 (34.4)
STD § (n = 271)	143 (52.8)	12 (8.4)
Cirrhosis¶ (n = 97)	87 (89.7)	45 (51.7)
ESRD** (n = 74)	63 (85.1)	6 (9.5)
Psychiatric diagnosis †† (n = 1550)	771 (49.7)	121 (15.7)
ALT elevation		
AnyALT > 40 U/L (n = 826)	555 (67.2)	169 (30.5)
All ALT ≤ 40 U/L (n = 8753)	3248 (37.1)	269 (8.3)
Combined Factors		
Any risk factor (n = 5262)	2559 (48.6)	401 (15.7)
No risk factor (n = 4317)	1244 (28.8)	37 (3.0)
Total (n = 9579)	3803 (39.7)	438 (11.5)

*ICD-9 or positive urine toxicology. †lCD-9 for Etoh dependence or etoh liver disease or etoh level ≥ 80. ‡ICD-9 or positive antibody test or western blot. §STD, Sexually Transmitted Disease (not HIV): ICD-9 or Positive GC or Chlamydia PCR probe. ¶ICD-9Code. **ESRD, End-Stage Renal Disease: ICD-9 code or procedure code for hemodialysis. ††ICD-9 for affective, anxiety, schizophrenia, or psychosis.

adjustment for birth-cohort status including substance abuse, HIV, cirrhosis, and ALT elevation. Lastly, in our study the risk-based screening strategy yielded high rates of anti-HCV positivity in all categories of risk in patients born outside the high-risk birth-cohort. These data suggest that current risk-based screening methods should be continued, and serious consideration should be given to expanding screening recommendations to include birth in the high-risk cohort. Birth cohort testing alone, however, is not recommended.

In this clinic population of an urban academic medical center, the conservative (floor) estimate of the prevalence of hepatitis C antibodies was 4.6%, almost three times the estimated national prevalence [1]. Our model designed to predict positivity in the untested population estimated a much higher overall prevalence, 7.7%, which is close to the prevalence of 8.3% reported in a similar population by McGinn [5]. Overall, 39.7% of subjects had been tested. Among those with identified risk (either born in the high prevalence birth-cohort, had a high-risk co-morbidity, or an elevated ALT level), 48.6% had been tested.

It is worth noting that the proportion tested was very high (28.8%) among patients with no identified risk (born outside the high prevalence birth-cohort, no high-risk co-morbidity, and no elevation of ALT) and that the rate of positivity in this group was substantial (3.0%), though less than those with identified risks. Whether a substantial proportion of these tested patients had risk factors not identified through the EMR is not clear. It is also possible that some patients without apparent risk were tested because patients or providers were responding to New York Department of Health efforts, begun in 2004, to raise Bronx community and provider awareness of HCV infection [39]. Because of the high underlying prevalence of HCV infection (between 4.6% and

Table 3 Factors associated with Hepatitis C testing

	Univariate		Multivariate	
	OR_{unadj}	95% CI	OR_{adj}	95% CI
In high-risk birth cohort*	1.64	1.51–1.78	1.39	1.27–1.52
Male	1.70	1.55–1.86	1.35	1.22–1.49
African American	1.08	0.99–1.18	1.22	1.06–1.39
Latino	1.03	0.95–1.11	1.16	1.03–1.32
Substance Abuse†	6.00	4.89–7.37	3.20	2.57–4.00
Alcohol Abuse‡	4.50	3.19–6.36	1.96	1.33–2.90
HIV§	11.59	8.69–15.47	7.75	5.75–10.43
STD¶	1.72	1.35–2.20	1.89	1.46–2.44
Cirrhosis**	13.50	7.01–26.01	4.65	2.30–9.41
ESRD††	8.83	4.65–16.77	8.99	4.68–17.28
Psychiatric Diagnosis‡‡	1.63	1.46–1.82	1.42	1.26–1.60
Any ALT > 40 U/L	3.47	2.98–4.04	2.63	2.24–3.09

*Born 1945–1964. †lCD-9 or positive urine toxicology. ‡lCD-9 for Etoh dependence or etoh liver disease or etoh level ≥ 80. §lCD-9 or positive antibody test or western blot. ¶STD, Sexually Transmitted Disease (not HIV): lCD-9 or Positive GC or Chlamydia PCR probe. **ICD-9 Code. ††ESRD, End-Stage Renal Disease: ICD-9 code or procedure code for hemodialysis. ‡‡ICD-9 for affective, anxiety, schizophrenia, or psychosis.

Table 4 Factors associated with Hepatitis C positivity in those tested

	Univariate		Multivariate	
	OR_{unadj}	95% CI	OR_{adj}	95% CI
In high-risk birth cohort*	3.78	3.03–4.72	2.73	2.14–3.49
Male	2.62	2.14–3.20	1.49	1.18–1.89
African American	0.88	0.71–1.10		–
Latino	1.18	0.96–1.44		–
Substance Abuse†	9.85	7.83–12.39	5.95	4.59–7.72
Alcohol Abuse‡	4.09	2.79–6.01		–
HIV§	5.29	4.15–6.75	3.07	2.30–4.10
STD¶	0.70	0.38–1.27		–
Cirrhosis**	9.06	5.87–13.97	4.24	2.51–7.18
ESRD††	0.81	0.35–1.88		–
Psychiatric Diagnosis‡‡	1.59	1.27–2.00		–
AnyALT > 40 U/L	4.85	3.89–6.04	3.75	2.90–4.84

*Born 1945–1964. †lCD-9 or positive urine toxicology. ‡lCD-9 for Etoh dependence or etoh liver disease or etoh level ≥ 80. §lCD-9 or positive antibody test or western blot. ¶STD, Sexually Transmitted Disease (not HIV): lCD-9 or Positive GC or Chlamydia PCR probe. **ICD-9 Code. ††ESRD, End-Stage Renal Disease: ICD-9 code or procedure code for hemodialysis. ‡‡ICD-9 for affective, anxiety, schizophrenia, or psychosis.

7.7%) in this population, universal testing for high-risk urban populations may be more appropriate than the risk-based screening strategy.

This analysis has several important limitations. First, not all patients were tested for anti-HCV so the prevalence we report is an estimate based on risk profile. Second, we utilized an EMR for data collection so we were unable to capture all risks for HCV infection for each patient. Lastly, we did not take into account the temporal relationship between risk factors and HCV tests. It is possible, for example, that a substance abuse diagnosis might have been coded after a HCV test was ordered, and thus we cannot be sure that the diagnosis of substance abuse was present, or in the physician's mind, at the time of testing. Despite these limitations,

we were able to uncover a strong relationship between high-risk co-morbidities and physician testing behavior.

In conclusion, we found a very high estimated prevalence of HCV infection in a high-risk urban patient population with a high prevalence of risk factors. We found strong evidence that physicians are using a risk-based screening strategy to identify patients with HCV infection, using known risk factors and other conditions associated with HCV to guide testing. We also found evidence that screening recommendations should be expanded to include the high prevalence birth cohort.

ACKNOWLEDGEMENTS

This project was funded by a Centers for Disease Control and Prevention contract via the Agency Health Care Research and Quality ACTION initiative to Boston University, contract HHSA2902006000012 T0#4, the Clinical Investigation Core of the Center for AIDS Research at the Albert Einstein College of Medicine and MMC, funded by the National Institutes of Health (NIH P30 AI51519), and the CTSA Grant UL1 RR025750 and KL2 RR025749 and TL1 RR025748 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH).

CONFLICTS OF INTEREST

Alain H. Litwin has served as a speaker for Roche Pharmaceuticals, and an advisory board member for Vertex Pharmaceuticals. Cindy L. Christiansen is receiving funding from Sanofi Aventis. Mari-Lynn Drainoni has been a consultant on research projects for DiMagi, Inc. Devin Thompson is an employee of Emerging Health Information Technologies. Bryce D. Smith and Cindy M. Weinbaum are employees of the Centers for Disease Control and Prevention.

REFERENCES

1 Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. *Ann Intern Med* 2006; 144(10): 705–714.
2 Campsmith ML, Rhodes P, Hall HI, Green T. HIV prevalence estimates – United States, 2006. *MMWR Morb Mortal Wkly Rep* 2008; 57(39): 1073–1076.
3 Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. Centers for Disease Control and Prevention. *MMWR Recomm Rep* 1998; 47(RR-19): 1–39.
4 El-Serag HB, Davila JA, Petersen NJ, McGlynn KA. The continuing increase in the incidence of hepatocellular carcinoma in the United States: an update. *Ann Intern Med* 2003; 139(10): 817–823.
5 McGinn T, O'Connor-Moore N, Alfandre D, Gardenier D, Wisnivesky J. Validation of a hepatitis C screening tool in primary care. *Arch Intern Med* 2008; 168(18): 2009–2013.
6 Manns MP, McHutchison JG, Gordon SC *et al.* Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* 2001; 358(9286): 958–965.
7 Fried MW, Shiffman ML, Reddy KR *et al.* Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002; 347(13): 975–982.
8 Hadziyannis SJ, Sette H Jr, Morgan TR *et al.* Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann Intern Med* 2004; 140(5): 346–355.
9 Strader DB, Wright T, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C. *Hepatology* 2004; 39(4): 1147–1171.
10 Ghany MG, Strader DB, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology* 2009; 49(4): 1335–1374.
11 Volk ML, Tocco R, Saini S, Lok AS. Public health impact of antiviral therapy for hepatitis C in the United States. *Hepatology* 2009; 50(6): 1750–1755.
12 Culver DH, Alter MJ, Mullan RJ, Margolis HS. Evaluation of the effectiveness of targeted lookback for HCV infection in the United States-interim results. *Transfusion* 2000; 40(10): 1176–1181.
13 Gordon FD. Cost-effectiveness of screening patients for hepatitis C. *Am J Med* 1999; 107(6B): 36S–40S.
14 Shin A, Cho ER, Kim J *et al.* Factors associated with awareness of infection status among chronic hepatitis B and C carriers in Korea. *Cancer Epidemiol Biomarkers Prev* 2009; 18(6): 1894–1898.
15 Dubois F, Desenclos JC, Mariotte N, Goudeau A. Hepatitis C in a French population-based survey, 1994: seroprevalence, frequency of viremia, genotype distribution, and risk factors. The Collaborative Study Group. *Hepatology* 1997; 25(6): 1490–1496.
16 EASL International Consensus Conference on hepatitis C. Paris, 26–27 February 1999. Consensus statement. *J Hepatol* 1999; 31(Suppl 1): 3–8.
17 Galmiche JP. French consensus conference on hepatitis C: screening and treatment. *Gut* 1998; 42(6): 892–898.
18 Lapane KL, Jakiche AF, Sugano D, Weng CS, Carey WD. Hepatitis C infection risk analysis: who should be screened? Comparison of multiple screening strategies based on the National Hepatitis Surveillance Program. *Am J Gastroenterol* 1998; 93(4): 591–596.
19 Singer ME, Younossi ZM. Cost effectiveness of screening for hepatitis C virus in asymptomatic, average-risk adults. *Am J Med* 2001; 111(8): 614–621.
20 Sroczynski G, Esteban E, Conrads-Frank A *et al.* Long-term effectiveness and cost-effectiveness of screening for hepatitis C virus infection. *Eur J Public Health* 2009; 19(3): 245–253.
21 Degos F. Hepatitis C and alcohol. *J Hepatol* 1999; 31(Suppl 1): 113–118.
22 Caldwell SH, Jeffers LJ, Ditomaso A *et al.* Antibody to hepatitis C is common among patients with alcoholic liver disease with and without risk factors. *Am J Gastroenterol* 1991; 86(9): 1219–1223.

23 Weinstock HS, Bolan G, Reingold AL, Polish LB. Hepatitis C virus infection among patients attending a clinic for sexually transmitted diseases. *JAMA* 1993; 269(3): 392–394.
24 Thomas DL, Zenilman JM, Alter HJ *et al.* Sexual transmission of hepatitis C virus among patients attending sexually transmitted diseases clinics in Baltimore – an analysis of 309 sex partnerships. *J Infect Dis* 1995; 171(4): 768–775.
25 Gunn RA, Murray PJ, Brennan CH, Callahan DB, Alter MJ, Margolis HS. Evaluation of screening criteria to identify persons with hepatitis C virus infection among sexually transmitted disease clinic clients: results from the San Diego Viral Hepatitis Integration Project. *Sex Transm Dis* 2003; 30(4): 340–344.
26 Kilbourne AM, McCarthy JF, Himelhoch S, Welsh D, Hauser P, Blow FC. Guideline-concordant hepatitis C virus testing and notification among patients with and without mental disorders. *Gen Hosp Psychiatry* 2008; 30(6): 495–500.
27 Himelhoch S, McCarthy JF, Ganoczy D *et al.* Understanding associations between serious mental illness and hepatitis C virus among veterans: a national multivariate analysis. *Psychosomatics* 2009; 50(1): 30–37.
28 Rosenberg SD, Goodman LA, Osher FC *et al.* Prevalence of HIV, hepatitis B, and hepatitis C in people with severe mental illness. *Am J Public Health* 2001; 91(1): 31–37.
29 Dinwiddie SH, Shicker L, Newman T. Prevalence of hepatitis C among psychiatric patients in the public sector. *Am J Psychiatry* 2003; 160(1): 172–174.
30 Alter MJ, Kruszon-Moran D, Nainan OV *et al.* The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med* 1999; 341(8): 556–562.
31 O'Brien SF, Fan W, Xi G *et al.* Declining hepatitis C rates in first-time blood donors: insight from surveillance and case-control risk factor studies. *Transfusion* 2008; 48(5): 902–909.
32 Bornschlegel K, Berger M, Garg RK *et al.* Prevalence of Hepatitis C Infection in New York City, 2004. *J Urban Health* 2009; 86(6): 909–917.
33 Hepatitis A, B, and C Surveillance Report. New York City Department Of Health and Mental Hygiene [2005 [cited 2009 Aug. 19]; Available from: URL:http://www.nyc.gov/html/doh/downloads/pdf/cd/cd-hepabc-surveillance-report.pdf
34 Public Welfare: Subtitle A—Department of Health and Human Services, General Administrative Requirements. http://www hhs gov/ocr/hipaa/finalreg html [2002 45 C.F.R. Part 160 Available from: URL:http://www.hhs.gov/ocr/hipaa/finalreg.html
35 Public Welfare: Subtitle A—Department of Health and Human Services, Security and Privacy. http://www hhs gov/ocr/hipaa/finalreg html [2002 (45 C.F.R. Part 164, Subparts A and E) Available from: URL:http://www.hhs.gov/ocr/hipaa/finalreg.html
36 Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality RMD. Clinical Classifications Software (CCS). 2007.
37 Prati D, Taioli E, Zanella A *et al.* Updated definitions of healthy ranges for serum alanine aminotransferase levels. *Ann Intern Med* 2002; 137(1): 1–10.
38 Pratt DS, Kaplan MM. Evaluation of abnormal liver-enzyme results in asymptomatic patients. *N Engl J Med* 2000; 342(17): 1266–1271.
39 Rude EJ, Weisfuse I. A community experience responding to hepatitis C. *Public Health Rep* 2007; 122(Suppl 2): 89–90.

Vascular Medicine
16(4) 253–259

DOI: 10.1177/1358863X11408640
vmj.sagepub.com

Relationship between central and peripheral atherosclerosis and left ventricular dysfunction in a community population

Connie W Tsao[1], Philimon Gona[2,3], Carol Salton[1], Joanne M Murabito[3,4], Noriko Oyama[1,a], Peter G Danias[1], Christopher J O'Donnell[3,5], Warren J Manning[1,6] and Susan B Yeon[1]

Abstract

We aimed to determine the relationships between resting left ventricular (LV) wall motion abnormalities (WMAs), aortic plaque, and peripheral artery disease (PAD) in a community cohort. A total of 1726 Framingham Heart Study Offspring Cohort participants (806 males, 65 ± 9 years) underwent cardiovascular magnetic resonance with quantification of aortic plaque volume and assessment of regional left ventricular systolic function. Claudication, lower extremity revascularization, and ankle–brachial index (ABI) were recorded at the most contemporaneous examination visit. WMAs were associated with greater aortic plaque burden, decreased ABI, and claudication in age- and sex-adjusted analyses (all $p < 0.001$), which were not significant after adjustment for cardiovascular risk factors. In age- and sex-adjusted analyses, both the presence ($p < 0.001$) and volume of aortic plaque were associated with decreased ABI ($p < 0.001$). After multivariable adjustment, an ABI ≤ 0.9 or prior revascularization was associated with a threefold odds of aortic plaque ($p = 0.0083$). Plaque volume significantly increased with decreasing ABI in multivariable-adjusted analyses ($p < 0.0001$). In this free-living population, associations of WMAs with aortic plaque burden and clinical measures of PAD were attenuated after adjustment for coronary heart disease risk factors. Aortic plaque volume and ABI remained strongly negatively correlated after multivariable adjustment. Our findings suggest that the association between coronary heart disease and non-coronary atherosclerosis is explained by cardiovascular risk factors. Aortic atherosclerosis and PAD remain strongly associated after multivariable adjustment, suggesting shared mechanisms beyond those captured by traditional risk factors.

Keywords

aortic atherosclerosis; epidemiology; left ventricular wall motion abnormality; MRI; peripheral artery disease

Introduction

Central and peripheral artery disease (PAD) are manifestations of cardiovascular disease (CVD) that carry significant morbidity and mortality. In prospective, population-based studies, the presence and extent of atherosclerotic plaques in the carotid and femoral arteries are associated with increased CVD death.[1,2] Decreased ankle–brachial index (ABI) is associated with increased risk of congestive heart failure (CHF), coronary heart disease (CHD), and CVD death,[3,4] and confers increased morbidity and mortality for each category of the Framingham Risk score (FRS).[5] CVD risk increases with greater extent of atherosclerosis across multiple vascular territories.[6]

Myocardial ischemia and infarction resulting from CHD is a major cause of resting left ventricular (LV) wall motion abnormalities (WMAs). Echocardiographic resting WMAs correlate with significant angiographic coronary stenoses.[7] In participants with known CHD and LV dysfunction, the presence and severity of resting WMAs are associated with increased morbidity and mortality.[8,9] In subjects without overt CVD, WMAs may be a marker of silent CHD.[10]

[1] Harvard-Thorndike Laboratory and Department of Medicine (Cardiovascular Division), Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA
[2] Department of Mathematics and Statistics, Boston University, Boston, MA, USA
[3] National Heart, Lung and Blood Institute of the National Institutes of Health's Framingham Heart Study, Framingham, MA, USA
[4] Department of Medicine, Section of General Internal Medicine, Boston University, Boston, MA, USA
[5] Department of Medicine (Cardiovascular Division), Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[6] Harvard-Thorndike Laboratory and Department of Radiology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA
[a] Present address: Department of Radiology, Hokkaido University Graduate School of Medicine, Sapporo, Japan

Corresponding author:
Susan B Yeon
Division of Cardiology
Beth Israel Deaconess Medical Center
330 Brookline Avenue
Boston, MA 02215
USA
Email: syeon@bidmc.harvard.edu

Cardiovascular magnetic resonance (CMR) facilitates both qualitative and quantitative measures of LV systolic function, as well as aortic plaque burden. While the extent of polyvascular atherosclerotic disease has been reported,[6] there are few studies that examine the inter-relatedness of WMAs, aortic atherosclerosis, and PAD in an unselected population, particularly including subjects without prevalent CVD. As atherosclerosis has a prolonged subclinical course and correlates with future adverse CVD events,[3,4] early intervention may prevent the development of overt CVD or recurrent events, provided that similar mechanisms produce atherosclerosis in different vascular beds. We hypothesized that measures of aortic and peripheral atherosclerosis would be associated both with WMAs and with each other in a free-living population.

Methods

Participants

Study participants included a subset of the 3539 participants attending Examination 7 (1998–2001) of the Framingham Heart Study (FHS) Offspring Cohort, which has been described previously.[11] At evaluation every 3–4 years, beginning with Examination 1 (1971–1975), participants underwent routine medical history and physical exam, anthropometry, and assessment of CVD risk factors. For the CMR substudy, a random sampling strategy was used to recruit from strata of sex, decade age, and quintile of the FRS.[12] Participants were excluded if they were not in sinus rhythm, had a contraindication to CMR (e.g. pacemaker), or did not live in a state contiguous with Massachusetts. CMR scanning (2002–2006) was incomplete in 32 participants (claustrophobia, $n = 13$; scanner dysfunction, $n = 7$; metallic devices, $n = 10$; miscellaneous, $n = 2$). A total of 1726 participants (aged 65 ± 9 years, 806 men) completed CMR with analyzable images. ABI data were available in 1678 of these participants. The study was approved by the institutional review boards of both the Boston University Medical Center and the Beth Israel Deaconess Medical Center. All participants provided written informed consent.

CMR imaging

Supine CMR imaging was performed using a 1.5 T CMR scanner (Gyroscan NT; Philips Medical Systems, Best, The Netherlands) with a five-element commercial cardiac array coil for radio frequency signal reception. Following localizing scans to determine the position and orientation of the heart within the thorax, end-expiratory breath-hold, ECG-gated cine steady-state free precession images were acquired in two-chamber, four-chamber, and contiguous short axis orientations (temporal resolution 39 ms, repetition time = R-R interval, echo time 9 ms, flip angle 30 degrees, field of view 400 mm, matrix size 208 × 256, slice thickness 10 mm, gap = 0).

Aortic plaque imaging included 36 transverse slices from the aortic arch to the aortoiliac bifurcation using a free-breathing, ECG-gated, fat-suppressed, black blood T2-weighted turbo spin-echo sequence[13] with 5-mm slice thickness and an in-plane spatial resolution of 1.03 × 0.64 mm (10-mm and 5-mm slice gaps for thoracic and abdominal aorta, respectively).

Image analysis of LV function and aortic plaque

Cine-CMR image analysis was performed using dedicated software (EasyVision 5.1; Philips Medical Systems) by a single observer (C.S.) blinded to all clinical data. LV wall motion was analyzed according to a 17-segment model.[14] WMAs noted by the observer were confirmed by two additional reviewers blinded to all clinical data (C.T., S.Y). The global and regional wall motion score was computed using a five-point scale (1 = normal, 2 = hypokinetic, 3 = akinetic, 4 = dyskinetic, 5 = aneurysm), with the normal sum of all segments scoring 17. The wall motion score index (WMSI) was calculated as the total wall motion score divided by the number of segments, with a WMSI ≥ 19/17 (WMSI > 1.12, WMA) considered abnormal (≥ 2 contiguous hypokinetic segments, and/or one akinetic or dyskinetic segment).[10] Quantitative measures of LV systolic function and mass (LVM) were obtained by manually tracing epicardial and endocardial LV borders at end-diastole and end-systole, as previously described.[11] LV end-diastolic volume (EDV) and end-systolic volume (ESV) were computed using the summation of discs method. The LV ejection fraction (LVEF) was computed by (EDV – ESV)/EDV. The LVM was determined by summing myocardial volume and multiplying by myocardial density (1.05 g/ml). The LVM was indexed to body surface area (BSA). The LVM index (LVMI), relative wall thickness (RWT), LVM/LVEDV, and LVEF were tabulated.

CMR images were analyzed with commercial software (MASS v 6.1; QT-MEDIS, NL, Leiden, The Netherlands) for descending thoracic and abdominal aortic atherosclerotic plaque by a single expert reviewer (N.O.) blinded to all clinical data.[13] Images analyzed were perpendicular to the aorta with > 50% of the inner circumference of the aortic wall visualized. Atherosclerotic plaque was defined as luminal protrusions of ≥ 1 mm in radial thickness[13] that were visually distinguished from the minimal residual blood signal of each plaque. By visually tracing the plaque border, the cross-sectional area of plaque was measured, and total plaque volume was calculated. Inter- and intra-reader replicate measurements were made to determine reproducibility.[13]

Clinical variables and assessment of PAD

Participants underwent routine physical examination, anthropometry, and laboratory assessment of CVD risk factors at Examination 7 (1998–2001). Resting systolic (SBP) and diastolic blood pressure (DBP) were measured in the right arm seated position. Plasma glucose, and total and high-density lipoprotein cholesterol were measured on morning samples after an 8-hour fast. Hypertension was defined as SBP ≥ 140 mmHg or DBP ≥ 90 mmHg or use of

Table 1. Raw clinical characteristics of the study population

	No WMAs $n = 1620$	WMAs $n = 106$	p-value
Risk factors and clinical events			
Age (years)	64.5 ± 9.0	67.7 ± 9.1	0.0005
Male (%)	45	78	< 0.0001
BMI (kg/m^2)	27.6 ± 5.3	29.3 ± 5.0	0.0037
Obesity (%)	32.3	46.2	0.0033
Diabetes (%)	8.8	24.5	< 0.0001
Systolic blood pressure (mmHg)	124.9 ± 17.6	128.6 ± 13.9	0.0364
Diastolic blood pressure (mmHg)	74.0 ± 9.4	75.1 ± 11.3	0.3253
Hypertension (%)	50.5	78.3	< 0.0001
Tobacco use (pack-years)	13.4 ± 19.3	29.4 ± 28.2	< 0.0001
Current or former cigarette smoker (%)	59.8	72.6	0.0085
Total cholesterol/HDL	4.0 ± 1.3	4.4 ± 1.3	0.0216
Dyslipidemia (%)	79.5	87.7	0.04
FRS	7.64 ± 4.06	9.95 ± 3.86	< 0.0001
CHD or CHF (%)	6.7%	47.2%	< 0.0001
CMR measurements			
WMSI	1.00 ± 0.01	1.56 ± 0.43	
LVMI (g/m^2)	54.2 ± 11.1	68.3 ± 14.3	< 0.0001
LVEDV (ml)	123 ± 29	162 ± 36	< 0.0001
LVEF (%)	68.1 ± 5.7	53.6 ± 9.1	< 0.0001
Aortic plaque measurements			
Aortic plaque, n (%)	771 (48)	60 (57)	0.072
Aortic plaque volume, cm^2 (IQR)	0.0 (0.0–0.4)	0.1 (0.0–1.2)	< 0.001
Clinical PAD[a]			
Intermittent claudication, n (%)	31 (2)	9 (9)	< 0.001
ABI, mean (SD)	1.14 ± 0.10	1.11 ± 0.15	< 0.001
ABI ≤ 0.9 or revascularization, n (%)	34 (2)	8 (8)	< 0.001
ABI 0.91–1.0, n (%)	68 (4)	8 (8)	< 0.001

Continuous summary measures are mean ± SD.
WMA, wall motion abnormalities; BMI, body mass index; HDL, high-density lipoprotein; FRS, Framingham Coronary Risk Score; CHD, myocardial infarction, angina pectoris, or coronary insufficiency; CHF, congestive heart failure; CMR, cardiovascular magnetic resonance; WMSI, wall motion score index; LVMI, left ventricular mass index; LVEDV, left ventricular end-diastolic volume; LVEF, left ventricular ejection fraction; IQR, interquartile range; PAD, peripheral artery disease; ABI, ankle–brachial index.
WMA = WMSI > 1.12; Obesity = BMI > 30; Dyslipidemia = total cholesterol ≥ 200 mg/dl or on lipid-lowering therapy; Tobacco = any smoking during Examinations 1–7.
[a]PAD data available in $n = 1671$.

antihypertensive medications. Dyslipidemia was defined as a total cholesterol ≥ 200 mg/dl or the use of lipid-lowering therapy. Diabetes mellitus was defined as fasting glucose ≥ 126 mg/dl or the use of insulin or oral hypoglycemic medications. A history of CHD (defined as recognized or unrecognized myocardial infarction [with diagnostic electrocardiography], angina, or coronary insufficiency) and CHF (identified by clinical signs and symptoms) were determined by physician end-point review.[12]

The presence of intermittent claudication at any clinic examination visit was defined by a physician-administered questionnaire, in which participants reported exertional leg discomfort related to degree of walking that was relieved with rest, verified by an endpoint review panel of three investigators.[15] Ankle–brachial SBP measurements, repeated twice, were obtained by trained technicians using a standard protocol at Examination 7 (average 4.4 years from CMR study).[15] ABI was calculated for each leg as the ratio of average SBP in the ankle divided by average SBP in the higher arm. The lower of the two ABIs calculated for each lower extremity was used for analysis. If ABI was missing for one lower extremity, data were used from the other extremity. A reported history of lower extremity revascularization, including percutaneous angioplasty, placement of stent, or vascular bypass surgery, was recorded and validated by medical record review. Three categories of ABI were defined: ABI ≤ 0.9 or history of lower extremity revascularization (significant PAD); 0.91 ≤ ABI ≤ 1.0; and 1.0 < ABI ≤ 1.4. Participants with ABI > 1.4 were excluded ($n = 7$) as these values may represent medial arterial calcification with associated increased mortality.[16] This exclusion resulted in $n = 1671$ participants with ABI data for analysis.

Statistical analysis

Participants were categorized by the presence or the absence of a WMA. Plaque volume was natural log-transformed due to a non-normal distribution and reported as median (interquartile range). Descriptive statistics for all covariates are presented as percentages or means ± SD. Differences in characteristics between the groups with and without WMAs were evaluated using two-sample t-tests and analysis of

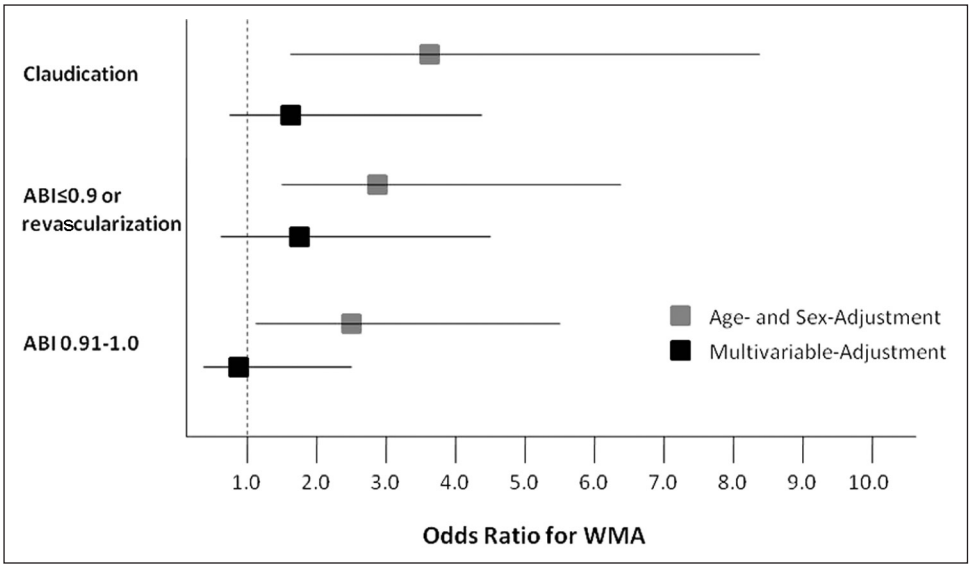


Figure 1. Odds of a WMA associated with peripheral artery disease (PAD). Boxes represent odds ratio for a WMA; lines represent 95% CI. Claudication, ABI ≤ 0.9 or revascularization, and ABI 0.91–1.0 were associated with increased odds of a WMA in age- and sex-adjusted analyses. These associations were attenuated after multivariable adjustment.

covariance (ANCOVA) for continuous variables, and the chi-squared test and logistic regression for binary variables. Age- and sex-adjusted and multivariable-adjusted ANCOVA and logistic regression models were constructed to assess the association of aortic plaque, natural log-transformed plaque volume, claudication, and ABI (as both continuous and categorical variables, using ABI > 1.0 to ≤ 1.4 as the referent group) with the presence or absence of a WMA. Odds ratios and 95% confidence intervals were calculated for the association of categorical variables with WMAs. Covariates in the multivariable model were determined at Examination 7: age, sex, BMI, tobacco pack-years, SBP, total cholesterol/high-density lipoprotein (HDL) ratio, and histories of hypertension treatment, dyslipidemia treatment, and diabetes mellitus. Similar models were generated to assess the relationship of aortic plaque to ABI severity. An age- and sex-adjusted Pearson correlation coefficient was used to determine the linear association between the presence and quantity of aortic plaque with ABI. All analyses were performed with SAS 8.0 (SAS Institute, Cary, NC, USA). A two-sided p-value < 0.05 was considered statistically significant.

The authors had full access to and take full responsibility for the integrity of the data. All authors have read and agree to the manuscript as written.

Results

The demographics and characteristics of the study sample are shown in Table 1. WMAs were present in 106 (6%) of the total population. Compared to participants without a WMA, those with WMAs had a greater prevalence of all cardiovascular risk factors, particularly male sex, diabetes, and hypertension, and a greater prevalence of CHD or CHF (all $p < 0.001$). However, 53% of participants with WMAs had no history of CHD or CHF. Participants with WMAs also had greater LVMI and LVEDV, and lower LVEF (all $p < 0.0001$).

The prevalence of aortic plaque and raw measures of clinical PAD are presented in Table 1. Aortic plaque was present in 48% of the cohort. The majority of participants with aortic plaque (93%) had normal LV wall motion. Plaque volume was greater in participants with WMAs compared to those with normal wall motion ($p < 0.001$). This difference persisted after age- and sex-adjustment ($p = 0.001$) but was attenuated after multivariable adjustment ($p = 0.22$).

Participants with WMAs had a greater prevalence of both claudication and decreased ABI. Claudication was associated with a nearly fourfold odds of having a WMA in age- and sex-adjusted analyses (OR 3.78, 95% CI 1.68–8.52, $p = 0.001$); this association was attenuated after multivariable adjustment (OR 1.69, 95% CI 0.66–4.37, $p = 0.276$). Mean ABI was not significantly lower in participants with a WMA in multivariable adjusted analyses (mean ± SD: 1.11 ± 0.15 vs 1.14 ± 0.10, $p = 0.12$). ABI ≤ 0.9 or revascularization were present in 42 (2.5%) participants, ABI 0.91–1.0 was present in 76 (4.5%) participants, and ABI > 1.0–1.4 was present in 1553 (92.9%) participants. There was a significant increase in the proportion of participants with WMAs across decreasing levels of ABI (5.5% among ABI > 1.0 to 1.4, 10.5% among ABI 0.91–1.0, and 19.0% among ABI ≤ 0.9 or revascularization, $p < 0.001$ for linear trend). In age- and sex-adjusted analyses, ABI ≤ 0.9 or revascularization was associated with a nearly threefold risk of WMA (OR 2.94, 95% CI 1.28–6.76, $p = 0.01$) and ABI 0.91–1.0 was associated with an elevated risk for WMAs (OR 2.58, 95% CI 1.13–5.84, $p = 0.02$) compared to ABI > 1.0 and ≤ 1.4. However, the association between ABI level and WMA was not significant after multivariable adjustment (OR for WMA 1.9, 95% CI 0.78–4.40, $p = 0.843$, for ABI ≤ 0.9 or revascularization; OR for WMA 0.90, 95% CI 0.32–2.52, $p = 0.163$, for ABI 0.91–1.0) (Figure 1).

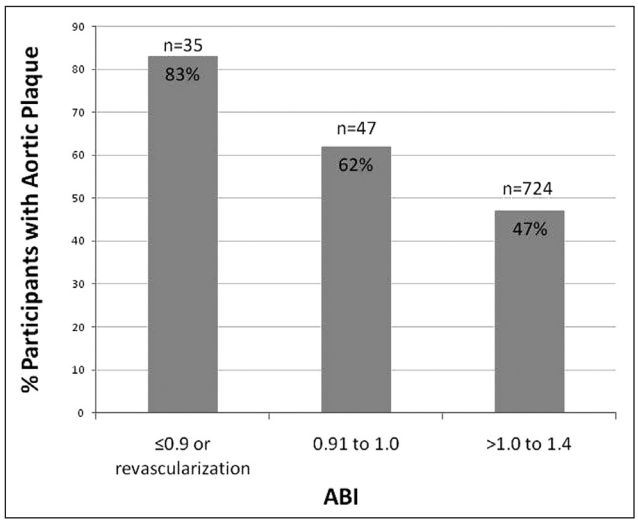


Figure 2. Relationship of aortic plaque prevalence with ABI groups. The prevalence of aortic plaque increased with decreasing ABI group ($p < 0.001$).

Aortic plaque and ABI were inversely associated. In participants with aortic plaque, ABI was significantly lower compared to those without plaque (ABI = 1.12 ± 0.11 vs 1.15 ± 0.09, $p < 0.0001$). The prevalence of aortic plaque increased with decreasing ABI category (Figure 2). ABI ≤ 0.9 or a history of revascularization was associated with fivefold increase in the odds of aortic plaque in age- and sex-adjusted analyses (OR 4.99, 95% CI 2.184–11.414, $p = 0.0014$) and a threefold increase in the odds after multivariable adjustment (OR 3.23, 95% CI 1.35–7.73, $p = 0.0083$). Likewise, median aortic plaque burden also increased with decreasing ABI group, from 0 cm^3 (IQR 0.0–0.39 cm^3) among ABI > 1.0 to 1.4, to 0.34 cm^3 (IQR 0.0–1.49 cm^3) among ABI 0.91–1.0, to 2.2 cm^3 (IQR 0.23–4.10 cm^3) among ABI ≤ 0.9 or with a history of revascularization. This relationship was significant in age- and sex-adjusted analyses (Spearman $r = –0.28$, $p = 0.001$), and persisted after multivariable adjustment ($p < 0.0001$).

Discussion

To our knowledge, this is the first population-based study to evaluate the relationship of LV WMAs, a likely surrogate for CHD, with both aortic and peripheral atherosclerosis. Notably, a significant proportion of participants with WMAs were free of prevalent CHD or CHF. While LV WMAs may result from non-ischemic etiologies, underlying CHD is a common etiology in middle-aged populations in developed countries.[17,18] The associations between extent of aortic plaque and clinical PAD with LV WMAs seen in age- and sex-adjusted analyses were not significant after multivariable adjustment. However, aortic plaque and ABI remained strongly inversely correlated after adjustment for CHD risk factors. While those with ABI data represented a smaller group, exclusion of these participants did not change the results of the prevalence and mean values of risk factors with the exception of obesity, whose lower prevalence was similar among participants with and without WMAs.

In our community-based population, CMR evidence of aortic plaque was common. Though many participants without WMAs had aortic plaque, this group had a lower overall plaque volume. The high prevalence of aortic plaque in both groups with and without WMAs is consistent with the high prevalence of CVD risk factors, notably hypertension, history of tobacco use, and dyslipidemia throughout the population. The association of increased plaque volume in participants with WMAs was attenuated in multivariable analysis, suggesting that these mechanisms of atherosclerosis in the aorta and coronary arteries have in common traditional CVD risk factors including obesity, smoking, diabetes, hypertension, and dyslipidemia. However, our results are consistent with transesophageal echocardiographic data showing an association of aortic plaque thickness with CHD[19] and an increased incidence of CHD and CVD morbidity in those with aortic calcification.[20–23]

Significant clinical PAD was uncommon in our cohort (2.4% with ABI ≤ 0.9 or revascularization and only 2.3% with claudication), despite the prevalence of CVD risk factors. This is similar to the 2–4% prevalence of ABI ≤ 0.9 reported in an earlier, larger subset of the FHS Offspring Cohort,[15] in the Atherosclerosis Risk in Communities (ARIC) Study,[24] in the Multi-Ethnic Study of Atherosclerosis (MESA),[25] and in a large Italian population.[26] The 8–9% prevalence of significant PAD among those with a WMA was greater than those with normal LV wall motion. The fact that ABI and WMAs were associated in age- and sex-adjusted analyses but not upon multivariable adjustment in our study suggests that both ABI and WMAs may be influenced by traditional CVD risk factors. A similar relationship was seen between claudication and WMAs in multivariable analyses.

While attenuated in multivariable analyses, the association between PAD and WMAs is in agreement with findings from the Cardiovascular Health Study, which reported an association between ABI < 0.9 and echocardiographic segmental WMAs and in age- and sex-adjusted analyses.[27] Our results are also consistent with that between decreased ABI and both prevalence and incident development of CHD in the ARIC studies.[24,28] While an ABI ≤ 0.9 is used as the cut-off for significant PAD,[5] an ABI 0.91–1.0 may reflect an intermediate risk, as this group still had a nearly threefold odds of WMA in age- and sex-adjusted analyses. These results are consistent with reports of both an increased prevalence of CHD[15,24] and both cohort and meta-analysis studies, demonstrating a greater risk for cardiovascular events and mortality with progressive decreases in ABI.[5,29]

Other studies have found a relationship between coronary artery calcification (CAC), another surrogate for CHD, and PAD. In a subset of 1932 participants without cardiovascular risk factors in MESA, a low ABI (< 1.0) was associated with the presence of CAC.[30] In addition, in a cohort of 3479 participants with CHD, those with concomitant PAD had a greater volume of coronary atheroma and extent of CAC.[31] Our study may show a lack of association in adjusted analyses due to the limited number of participants with PAD or due to the indirectness of WMAs as a marker of coronary artery disease as compared to CAC.

In contrast, both the prevalence and burden of aortic plaque demonstrated a strong inverse relationship with groups of ABI independent of traditional CVD risk factors. The stronger association of measures of non-coronary atherosclerosis with each other than with WMAs could reflect that WMAs are the result of a clinical event which may not correlate with absolute degree of atherosclerosis. This is consistent with past angiographic evidence that often angiographic severity may result from lesser grade, rather than highly stenotic lesions.[32] Alternatively, factors promoting atherosclerosis in coronary as compared with non-coronary beds may have different mechanisms. Consistent with this possibility, members of the Reduction of Atherothrombosis for Continued Health (REACH) cohort with PAD had a markedly greater prevalence of polyvascular disease than did those with CAD (60% vs 25%, respectively).[6]

CMR is advantageous as it is not limited by acoustic windows, generates standard imaging planes, and defines wall motion at the apex with higher precision than echocardiography.[33] We were able to assess LV wall motion in 99% of all participants and complete aortic plaque analysis in 96% of participants. This technique has been well-validated in the assessment of LV function[34] and imaging atherosclerotic plaque.[13]

Limitations of this study include a relatively small number of participants with clinical PAD and WMAs, which may have limited the power to detect associations between PAD and WMAs. In addition, Examination 7 data and CMR data were not obtained concurrently. While there was an average of 4.4 years between Examination 7 and the CMR test, this time interval would be more likely to bias true associations towards the null. Furthermore, assessment of regional WMAs remains largely observer-dependent, though CMR allows superior endocardial border definition and thus excellent interobserver agreement.[35] While the majority of WMAs (94%) reflected regional rather than global LV dysfunction, 6% of WMAs were global. This minority of WMAs could represent non-ischemic cardiomyopathy with no significant coronary artery disease. Finally, the FHS Offspring Cohort is a predominantly middle-aged and older Caucasian population. Thus, our results may not be generalizable to other races or ethnicities or age groups.

In conclusion, in this community-based population, incidentally detected WMAs which likely reflect CHD were associated with measures of aortic and peripheral atherosclerosis in age- and sex-adjusted, but not multivariable-adjusted models. The prevalence and burden of aortic plaque were strongly and independently associated with peripheral artery disease in both age- and sex- and multivariable-adjusted analyses. WMAs and peripheral atherosclerosis share common CVD risk factors. However, aortic atherosclerosis and PAD remain strongly associated after multivariable adjustment, suggesting shared mechanisms beyond those captured by traditional risk factors. While screening for WMAs is not advocated in a CHD at-risk population, unsuspected WMAs may be identified during imaging. Further prospective studies are necessary to determine the clinical utility of detection of WMAs relative to other markers of CHD.

Disclosures

The Framingham Heart Study is supported by Grant N01-HC-25195 from the National Heart, Lung and Blood Institute, Bethesda, Maryland, USA. This work is also partially supported by a grant from the National Institutes of Health (RO1 HL70279) and a National Institutes of Health training grant (T32 HL07374 to CWT).

References

1. Lamina C, Meisinger C, Heid IM, et al. Association of ankle–brachial index and plaques in the carotid and femoral arteries with cardiovascular events and total mortality in a population-based study with 13 years of follow-up. *Eur Heart J* 2006; 27: 2580–2587.
2. Stork S, van den Beld AW, von Schacky C, et al. Carotid artery plaque burden, stiffness, and mortality risk in elderly men: a prospective, population-based cohort study. *Circulation* 2004; 110: 344–348.
3. Criqui MH, Langer RD, Fronek A, et al. Mortality over a period of 10 years in patients with peripheral arterial disease. *N Engl J Med* 1992; 326: 381–386.
4. Newman AB, Shemanski L, Manolio TA, et al. Ankle–arm index as a predictor of cardiovascular disease and mortality in the Cardiovascular Health Study. The Cardiovascular Health Study Group. *Arterioscler Thromb Vasc Biol* 1999; 19: 538–545.
5. Fowkes FG, Murray GD, Butcher I, et al. Ankle brachial index combined with Framingham Risk Score to predict cardiovascular events and mortality: a meta-analysis. *JAMA* 2008; 300: 197–208.
6. Steg PG, Bhatt DL, Wilson PW, et al. One-year cardiovascular event rates in outpatients with atherothrombosis. *JAMA* 2007; 297: 1197–1206.
7. Whellan DJ, Tuttle RH, Velazquez EJ, et al. Predicting significant coronary artery disease in patients with left ventricular dysfunction. *Am Heart J* 2006; 152: 340–347.
8. Fauchier L, Eder V, Casset-Senon D, et al. Segmental wall motion abnormalities in idiopathic dilated cardiomyopathy and their effect on prognosis. *Am J Cardiol* 2004; 93: 1504–1509.
9. Shiina A, Tajik AJ, Smith HC, Lengyel M, Seward JB. Prognostic significance of regional wall motion abnormality in patients with prior myocardial infarction: a prospective correlative study of two-dimensional echocardiography and angiography. *Mayo Clin Proc* 1986; 61: 254–262.
10. Cicala S, de Simone G, Roman MJ, et al. Prevalence and prognostic significance of wall-motion abnormalities in adults without clinically recognized cardiovascular disease: the Strong Heart Study. *Circulation* 2007; 116: 143–150.
11. Salton CJ, Chuang ML, O'Donnell CJ, et al. Gender differences and normal left ventricular anatomy in an adult population free of hypertension. A cardiovascular magnetic resonance study of the Framingham Heart Study Offspring cohort. *J Am Coll Cardiol* 2002; 39: 1055–1060.
12. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998; 97: 1837–1847.

13. Oyama N, Gona P, Salton CJ, et al. Differential impact of age, sex, and hypertension on aortic atherosclerosis: the Framingham Heart Study. *Arterioscler Thromb Vasc Biol* 2008; 28: 155–159.
14. Cerqueira MD, Weissman NJ, Dilsizian V, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation* 2002; 105: 539–542.
15. Murabito JM, Evans JC, Nieto K, Larson MG, Levy D, Wilson PW. Prevalence and clinical correlates of peripheral arterial disease in the Framingham Offspring Study. *Am Heart J* 2002; 143: 961–965.
16. Resnick HE, Lindsay RS, McDermott MM, et al. Relationship of high and low ankle brachial index to all-cause and cardiovascular disease mortality: the Strong Heart Study. *Circulation* 2004; 109: 733–739.
17. Lambert CR Jr, Conti CR, Pepine CJ. Left ventricular function abnormalities as a manifestation of silent myocardial ischemia. *Cardiol Clin* 1986; 4: 669–675.
18. Gardin JM, Siscovick D, Anton-Culver H, et al. Sex, age, and disease affect echocardiographic left ventricular mass and systolic function in the free-living elderly. The Cardiovascular Health Study. *Circulation* 1995; 91: 1739–1748.
19. Tribouilloy C, Shen WF, Peltier M, Lesbre JP. Noninvasive prediction of coronary artery disease by transesophageal echocardiographic detection of thoracic aortic plaque in valvular heart disease. *Am J Cardiol* 1994; 74: 258–260.
20. Eisen A, Tenenbaum A, Koren-Morag N, et al. Calcification of the thoracic aorta as detected by spiral computed tomography among stable angina pectoris patients: association with cardiovascular events and death. *Circulation* 2008; 118: 1328–1334.
21. Iribarren C, Sidney S, Sternfeld B, Browner WS. Calcification of the aortic arch: risk factors and association with coronary heart disease, stroke, and peripheral vascular disease. *JAMA* 2000; 283: 2810–2815.
22. Wilson PW, Kauppila LI, O'Donnell CJ, et al. Abdominal aortic calcific deposits are an important predictor of vascular morbidity and mortality. *Circulation* 2001; 103: 1529–1534.
23. Witteman JC, Kannel WB, Wolf PA, et al. Aortic calcified plaques and cardiovascular disease (the Framingham Study). *Am J Cardiol* 1990; 66: 1060–1064.
24. Zheng ZJ, Sharrett AR, Chambless LE, et al. Associations of ankle–brachial index with clinical coronary heart disease, stroke and preclinical carotid and popliteal atherosclerosis: the Atherosclerosis Risk in Communities (ARIC) Study. *Atherosclerosis* 1997; 131: 115–125.
25. McDermott MM, Liu K, Criqui MH, et al. Ankle–brachial index and subclinical cardiac and carotid disease: the multi-ethnic study of atherosclerosis. *Am J Epidemiol* 2005; 162: 33–41.
26. Santo Signorelli S, Anzaldi M, Fiore V, et al. Study on unrecognized peripheral arterial disease (PAD) by ankle/brachial index and arterial comorbidity in Catania, Sicily, Italy. *Angiology* 2010; 61: 524–529.
27. Newman AB, Siscovick DS, Manolio TA, et al. Ankle–arm index as a marker of atherosclerosis in the Cardiovascular Health Study. Cardiovascular Heart Study (CHS) Collaborative Research Group. *Circulation* 1993; 88: 837–845.
28. Weatherley BD, Nelson JJ, Heiss G, et al. The association of the ankle–brachial index with incident coronary heart disease: the Atherosclerosis Risk In Communities (ARIC) study, 1987–2001. *BMC Cardiovasc Disord* 2007; 7: 3.
29. O'Hare AM, Katz R, Shlipak MG, Cushman M, Newman AB. Mortality and cardiovascular risk across the ankle–arm index spectrum: results from the Cardiovascular Health Study. *Circulation* 2006; 113: 388–393.
30. Aboyans V, McClelland RL, Allison MA, et al. Lower extremity peripheral artery disease in the absence of traditional risk factors. The Multi-Ethnic Study of Atherosclerosis. *Atherosclerosis* 2011; 214: 169–173.
31. Hussein AA, Uno K, Wolski K, et al. Peripheral arterial disease and progression of coronary atherosclerosis. *J Am Coll Cardiol* 2011; 57: 1220–1225.
32. Little WC, Constantinescu M, Applegate RJ, et al. Can coronary angiography predict the site of a subsequent myocardial infarction in patients with mild-to-moderate coronary artery disease? *Circulation* 1988; 78: 1157–1166.
33. Moon JC, Fisher NG, McKenna WJ, Pennell DJ. Detection of apical hypertrophic cardiomyopathy by cardiovascular magnetic resonance in patients with non-diagnostic echocardiography. *Heart* 2004; 90: 645–649.
34. Keenan NG, Pennell DJ. CMR of ventricular function. *Echocardiography* 2007; 24: 185–193.
35. Syed MA, Paterson DI, Ingkanisorn WP, et al. Reproducibility and inter-observer variability of dobutamine stress CMR in patients with severe coronary disease: implications for clinical research. *J Cardiovasc Magn Reson* 2005; 7: 763–768.

Journal of Addictive Diseases, 30:91–97, 2011

ISSN: 1055-0887 print / 1545-0848 online
DOI: 10.1080/10550887.2011.554775

Routledge
Taylor & Francis Group

Chronic Pain and Hepatitis C Virus Infection in Opioid Dependent Injection Drug Users

Judith I. Tsui, MD, MPH
Debra S. Herman, PhD
Malyna Kettavong, BA
Bradley J. Anderson, PhD
Michael D. Stein, MD

ABSTRACT. It is unknown whether infection with hepatitis C is a risk factor for pain among people who have used injection drugs. Multivariate regression was used to determine whether hepatitis C was associated with greater likelihood of reporting significant chronic pain and discomfort intolerance in a cohort of 97 injection drug users dependent on opioids. Study results suggest that participants with hepatitis C may be more likely to experience chronic pain (aOR = 1.98; 95% confidence interval = 0.76 to 5.12, $p = 0.16$). Furthermore, hepatitis C was found to be associated with a higher discomfort intolerance scale score, reflecting intolerance to physical discomfort ($\beta = 2.34$; 95% confidence interval = 0.06 to 4.62; $p = 0.04$). Hepatitis C may be a cause for chronic pain and discomfort intolerance that is overlooked among injection drug users dependent on opioids.

KEYWORDS. Hepatitis C virus, injection drug use, chronic pain, pain hypersensitivity

BACKGROUND

Chronic pain and substance use frequently coexist for reasons that are poorly understood. Individuals dependent on opioids appear particularly vulnerable to chronic pain and pain hypersensitivity. There is a high prevalence of chronic severe pain (37% to 61%) among individuals dependent on opioids who receive methadone maintenance,[1–3] and studies of individuals with human immunodeficiency virus (HIV) have found that those with a history of injection drug use report more pain than those with no injection drug use.[4–8] Furthermore, experimental pain studies have demonstrated lower pain thresholds among individuals dependent on opioids.[9–12]

Individuals who inject opioids are at a greater risk for hepatitis C infections.[13–16] In a large case series, hepatitis C was linked to painful conditions such as peripheral neuropathies and arthritis.[17] Furthermore, quality of life studies have observed greater bodily pain among hepatitis C positive patients than hepatitis C negative patients.[18] These studies suggest a high prevalence of painful conditions among

Judith I. Tsui is affiliated with Boston University School of Medicine, Boston Medical Center, Boston, MA. Debra S. Herman, Malyna Kettavong, Bradley J. Anderson, and Michael D. Stein are affiliated with Butler Hospital, and Warren Alpert School of Medicine at Brown University, Providence, RI.

Address correspondence to: Judith I. Tsui, MD, MPH, 801 Massachusetts Avenue, Room 2078, Boston, MA 02118 (E-mail: Judith.tsui@bmc.org).

This study was funded by the National Institute of Drug Abuse DA 022207, Clinical Trial # NCT 00475878. Dr. Stein is a recipient of a NIDA Mid-Career Award DA 000512. Dr. Tsui is a recipient of a NIDA Career Development Award K23 DA027367.

hepatitis C positive patients; however, support for this association has been limited by a lack of control patients or adjustment for important confounders, most notably substance use disorders.[19–21] Researchers have hypothesized that increased levels of inflammatory cytokines may provide a biologic link between hepatitis C and fibromyalgia,[22] but a large population-based study failed to support such an association.[23]

Although the clinical overlap of substance use, hepatitis C, and pain is commonly seen in medical settings, there has been little attention to hepatitis C as a direct etiology for pain. One study of veterans seeking addiction services found that hepatitis C seropositivity was associated with a three-fold increased risk for persistent pain.[24] Hepatitis C may be an underrecognized and potentially modifiable cause of pain for patients with substance use. This study was undertaken to examine whether reported infection with hepatitis C was associated with chronic pain and discomfort intolerance in a sample of injection drug users dependent on opioids who exhibited symptoms of depression.

METHODS

Study Sample and Design

This cross-sectional study used baseline data from participants in a randomized, double-blind, placebo-controlled trial to determine whether treatment for symptoms of depression increases treatment retention among patients dependent on opioids initiating office-based buprenorphine treatment. Study inclusion criteria included age between 18 and 65, a *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) diagnosis of opioid dependence, a score on the Modified Hamilton Depression Revised Scale[25] greater than 14, the absence of significant suicidal ideation, willingness and ability to complete a 3-month treatment with buprenorphine, no history of severe mental illness (bipolar disorder, schizophrenia, schizo-affective disorder, or paranoid disorder), no currently prescribed psychotropic or antidepressant medications or medications that could cause depression (including interferons), and the ability to complete the study assessment in English. Between November 2006 and May 2009, 932 individuals were screened by telephone, and of those, 394 callers appeared eligible for the study and were invited for an in-person screening visit. Of the 226 who attended this visit, 147 fully met criteria and agreed to enroll in the parent study. This study restricted analyses to 97 participants who were current or former injection drug users (i.e., reported ever injecting drugs as assessed by an HIV-risk screening instrument).[26]

Pain Outcomes

Pain severity was assessed through the use of the Visual Analogue Scale, a rating scale that allows participants to mark their pain on a 100-mm line that ranges from "no pain" to "worst pain imaginable" and translates their pain to a point value between 0 and 100.[27] Pain interference was assessed using the mean of the 7-item subscale from the Brief Pain Inventory (BPI) short form.[28] The subscale measures pain interference in different domains such as sleep, work, and relationships, rating each item from "0" (pain does not interfere) to "10" (pain completely interferes). The primary outcome was significant chronic pain, which was defined as pain that was of at least moderate in intensity (Visual Analogue Scale score $\geq$40) or caused at least moderate interference (Brief Pain Inventory interference score $\geq$5) and which had been present for at least 6 months. This definition was modified from that used in a study by Rosenblum et al.[3] The secondary outcome of interest was discomfort intolerance, which was measured using the Discomfort Intolerance Scale (DIS).[29] It is a 5-item scale designed to measure the construct of discomfort intolerance, which is defined as the individual's perception of his or her ability to tolerate uncomfortable sensations. The scale is comprised of 5 questions (sample questions: "I am more sensitive to feeling physical discomfort compared to most people" and "I take extreme measures to avoid feeling physically uncomfortable") that have a range of seven possible responses (0 = *Not at all like me* to 6 = *Extremely like me*). Responses were analyzed as a total summed score, with higher

responses indicating a lower tolerance for physical discomfort. Additional analyses used data on the reported location of pain participants had in the previous week, which included headache, abdominal pain, back pain, joint pain, or muscle pain.

Predictors

The primary predictor of interest was chronic hepatitis C infection defined as a positive response to the question: "Do you have hepatitis C virus (hepatitis C)?" Additional covariates were age, sex, race (White vs. non-White), severe depression (score greater than 28 on the Beck Depression Inventory II[30]), and report of starting opioids for pain. Covariates were selected on the basis of face validity (demographics) or known associations with pain (depression).[31] In addition, models were adjusted for self-report of starting opioids for pain because individuals with preexisting pain might use prescription opioids preferentially over injecting heroin, and thus be less likely to be hepatitis C positive. Initiation of opioids for pain was defined as positive response to the question: "Do you believe that you started using your primary opiate of addiction to relieve physical pain?"

Statistical Analysis

Analyses were performed using baseline study data. We examined differences in demographic and clinical variables between participants with and without reported hepatitis C using t tests and chi-squared tests. Multivariate logistic and linear regression were performed to determine the adjusted relative odds for significant chronic pain and the mean difference in DIS score associated with being infected with hepatitis C, respectively. Given the small sample size, a stepwise backward selection strategy was used to select a final parsimonious model using a p value < 0.2 or a $>10\%$ change in the hepatitis C coefficient as criteria for retention of covariates. Finally, prevalence of specific pain locations or types (i.e., headache, abdominal pain, back pain, joint pain, and muscle pain) were compared using chi-square tests. All statistical analyses were conducted using Stata version 10.0 (StatCorp, College Station, TX).

RESULTS

Of the 97 participants, 37 (38%) reported that they were infected with hepatitis C; none reported being infected with HIV. Participants with hepatitis C were slightly older than those who were not infected with hepatitis C (mean age = 42 vs. 37 years, respectively, $p < 0.01$), otherwise there were no statistically significant differences between the two groups in the distribution of other variables (Table 1). The prevalence of significant chronic pain was 33% in the study cohort. A higher prevalence of significant chronic pain was observed among individuals who self-reported being hepatitis C positive (14 of 37; 38%) compared to those who were hepatitis C negative (18 or 60; 30%), although this difference did not reach statistical significance ($p = 0.43$). The mean DIS score was higher (indicating greater discomfort intolerance) among hepatitis C positive individuals

TABLE 1. Characteristics of Opioid Dependent Injection Drug Users With and Without Hepatitis C Virus

Characteristics	Without Hepatitis C Virus (n = 60) No. (%)	With Hepatitis C Virus (n = 37) No. (%)	p value
Age (mean ± SD)	37 (± 10)	42 (± 9)	<0.01
Female	11 (18)	6 (16)	0.79
Non-White	13 (22)	8 (22)	0.99
Severe depression	27 (45)	18 (49)	0.73
Started opioids for pain	19 (32)	7 (19)	0.17

SD = standard deviation.

TABLE 2. Relative Odds for Significant Chronic Pain Associated With Hepatitis Adjusted C Virus in Opioid Dependent Injection Drug Users ($N = 97$)

	Full Model			Final Model		
	OR	95% CI	p Value	OR	95% CI	p value
Hepatitis C Virus	1.98	(0.72 to 5.43)	0.18	1.98	(0.76 to 5.12)	0.16
Age	1	(0.95 to 1.05)	0.94			
Female	0.75	(0.21 to 2.68)	0.66			
Non-White	1.26	(0.43 to 3.69)	0.67			
Severe depression	2.31	(0.87 to 6.19)	0.09	2.35	(0.89 to 6.24)	0.09
Initiated opioids for pain	4.31	(1.49 to 12.50)	<0.01	4.5	(1.6 to 12.67)	< 0.01

OR = odds ratio; CI = confidence interval.

than hepatitis C negative individuals (19 vs. 16, repectively; $p = 0.05$). After adjusting for other covariates, participants with hepatitis C appeared to have a nearly two-fold increased risk for significant chronic pain, although results did not reach statistical significance at the $p < 0.05$ level (Table 2). Participants with hepatitis C scored significantly higher on the DIS scale, indicating a greater intolerance to discomfort (Table 3). Among those who endorsed chronic pain, there were no statistically significant differences at the $p < 0.05$ level between reported hepatitis C positive and negative participants in the prevalence of pain in the following locations: headache (29% vs. 17%, respectively, $p = 0.42$), abdominal pain (14% vs. 0%, respectively, $p = 0.10$), back pain (64% vs. 61%, respectively, $p = 0.85$), joint pain (50% vs. 56%, respectively, $p = 0.76$), and muscle pain (36% vs. 28%, respectively, $p = 0.63$).

DISCUSSION

Among a cohort of injection drug users dependent on opioids, we observed that reported hepatitis C infection was associated with intolerance for discomfort. In addition, there was a moderately strong (albeit non-significant) association between hepatitis C and significant chronic pain. Although patients with hepatitis C have been observed to have a high prevalence of painful conditions,[20–22] this is the first study to find an independent association between hepatitis C and pain in a cohort of injection drug use.

Hepatitis C may be associated with chronic pain in injection drug use through several different mechanisms. First, it may result in hepatic and non-hepatic complications that cause pain.[17] Our study observed the prevalence of pain to be higher in multiple domains (headache,

TABLE 3. Adjusted Mean Difference in DIS Score Associated With Hepatitis C Virus in Opioid Dependent Injection Drug Users ($N = 97$)

	Full Model			Final Model		
	β	95% CI	p value	β	95% CI	p value
Hepatitis C Virus	2.31	(−0.19 to 4.80)	0.07	2.34	(0.06 to 4.62)	0.04
Age	−0.02	(−0.14 to 0.11)	0.77			
Female	2.23	(−0.79 to 5.25)	0.15	2.38	(-0.54 to 5.29)	0.11
Non-White	−0.26	(−3.01 to 2.49)	0.85			
Severe depression	0.49	(−1.89 to 2.86)	0.68			
Initiated opioids for pain	−0.56	(−3.15 to 2.04)	0.67			

CI = confidence interval.

abdominal pain, and muscle pain) among participants with hepatitis C. Second, hepatitis C is associated with depression,[32] which is a strong risk factor for pain.[31] However, because the association between hepatitis C and chronic pain was observed even after adjustment for severe depression (and the sample itself was restricted to injection drug use with depressive symptoms), this appears not to be the sole mechanism. Finally, it is possible that hepatitis C may cause pain intolerance and hypersensitivity through a cytokine-mediated pathway.[22,24] Inflammatory cytokines such as Tumor Necrosis Factor-α (TNF-α) and interleukin-6 have recently been implicated in the pathogenesis of centrally mediated pain,[33,34] and patients with hepatitis C have been observed to have increased levels of these cytokines.[35–41] Our finding that participants with hepatitis C reported greater intolerance to physical discomfort may provide some indirect support for this hypothesis, although further experimental studies on pain tolerance thresholds are needed.

This study has major limitations, and results should be considered exploratory in nature. There was likely misclassification of our main predictor because hepatitis C status was based on self-report rather than hepatitis C antibody or viral load testing. Data from a prior research study suggest that self-report of hepatitis C has good specificity (88%) but low sensitivity (77%).[42] However, the relatively low hepatitis C prevalence among individuals with a history of injection drug use (38%) suggests that there was substantial underreporting. This misclassification, if non-differential with regards to pain, should bias to the null, which would give greater strength to our findings. However, it is also possible that patients with pain might be more routinely screened for hepatitis C (because of increased health service use), which could bias our results. Finally, the cross-sectional nature of this study precludes inferences on causality. It is possible that seeking pain relief could cause riskier injecting behaviors that led to hepatitis C infection rather than the opposite (i.e., hepatitis C leading to pain). However, we are not aware of any prior research suggesting an association between pain and riskier injecting behaviors.

This study provides preliminary results suggesting that hepatitis C is associated with chronic pain and intolerance to physical discomfort in injection drug users dependent on opioids. Although the findings of this study are preliminary, they have important implications for clinical practice because chronic hepatitis C may be an overlooked and potentially treatable cause for pain among current and former injection drug use. More research is needed to understand how chronic viral infections such as hepatitis C affect risk for chronic pain in substance users to inform future interventions.

REFERENCES

1. Barry DT, Beitel M, Garnet B, Joshi D, Rosenblum A, Schottenfeld RS. Relations among psychopathology, substance use, and physical pain experiences in methadone-maintained patients. J Clin Psychiatry 2009; 70: 1213–8.

2. Jamison RN, Kauffman J, Katz NP. Characteristics of methadone maintenance patients with chronic pain. J Pain Symptom Manage 2000; 19:53–62.

3. Rosenblum A, Joseph H, Fong C, Kipnis S, Cleland C, Portenoy RK. Prevalence and characteristics of chronic pain among chemically dependent patients in methadone maintenance and residential treatment facilities. JAMA 2003; 289:2370–8.

4. Del Borgo C, Izzi I, Chiarotti F, Del Forno A, Moscati AM, Cornacchione E, Fantoni M. Multidimensional aspects of pain in HIV–infected individuals. AIDS Patient Care STDS 2001; 15:95–102.

5. Dobalian A, Tsao JC, Duncan RP. Pain and the use of outpatient services among persons with HIV: results from a nationally representative survey. Med Care 2004; 42:129–38.

6. Martin C, Pehrsson P, Osterberg A, Sonnerborg A, Hansson P. Pain in ambulatory HIV-infected patients with and without intravenous drug use. Eur J Pain 1999; 3:157–164.

7. Tsao JC, Dobalian A, Naliboff BD. Panic disorder and pain in a national sample of persons living with HIV. Pain 2004; 109:172–80.

8. Vogl D, Rosenfeld B, Breitbart W, Thaler H, Passik S, McDonald M, Portenoy RK. Symptom prevalence, characteristics, and distress in AIDS outpatients. J Pain Symptom Manage 1999; 18:253–62.

9. Compton P, Charuvastra VC, Kintaudi K, Ling W. Pain responses in methadone-maintained opioid abusers. J Pain Symptom Manage 2000; 20:237–45.

10. Compton P, Charuvastra VC, Ling W. Pain intolerance in opioid-maintained former opiate addicts: effect

of long-acting maintenance agent. Drug Alcohol Depend 2001; 63:139–46.

11. Doverty M, White JM, Somogyi AA, Bochner F, Ali R, Ling W. Hyperalgesic responses in methadone maintenance patients. Pain 2001; 90:91–6.

12. Pud D, Cohen D, Lawental E, Eisenberg E. Opioids and abnormal pain perception: new evidence from a study of chronic opioid addicts and healthy subjects. Drug Alcohol Depend 2006; 82:218–23.

13. Garfein RS, Vlahov D, Galai N, Doherty MC, Nelson KE. Viral infections in short-term injection drug users: the prevalence of the hepatitis C, hepatitis B, human immunodeficiency, and human T-lymphotropic viruses. Am J Public Health 1996; 86:655–61.

14. Lorvick J, Kral AH, Seal K, Gee L, Edlin BR. Prevalence and duration of hepatitis C among injection drug users in San Francisco, Calif. Am J Public Health 2001; 91: 46–7.

15. Murrill CS, Weeks H, Castrucci BC, Weinstock HS, Bell BP, Spruill C, Gwinn M. Age-specific seroprevalence of HIV, hepatitis B virus, and hepatitis C virus infection among injection drug users admitted to drug treatment in 6 US cities. Am J Public Health 2002; 92:385–7.

16. Thomas DL, Vlahov D, Solomon L, Cohn S, Taylor E, Garfein R, Nelson KE. Correlates of hepatitis C virus infections among injection drug users. Medicine (Baltimore) 1995; 74:212–20.

17. Cacoub P, Renou C, Rosenthal E, Cohen P, Loury I, Loustaud-Ratti V, Yamamoto AM, Camproux AC, Hausfater P, Musset L, Veyssier P, Raguin G, Piette JC. Extrahepatic manifestations associated with hepatitis C virus infection. A prospective multicenter study of 321 patients. The GERMIVIC. Groupe d'Etude et de Recherche en Medecine Interne et Maladies Infectieuses sur le Virus de l'Hepatite C. Medicine (Baltimore) 2000; 79:47–56.

18. Spiegel BM, Younossi ZM, Hays RD, Revicki D, Robbins S, Kanwal F. Impact of hepatitis C on health related quality of life: a systematic review and quantitative assessment. Hepatology 2005; 41:790–800.

19. Barkhuizen A, Rosen HR, Wolf S, Flora K, Benner K, Bennett RM. Musculoskeletal pain and fatigue are associated with chronic hepatitis C: a report of 239 hepatology clinic patients. Am J Gastroenterol 1999; 94:1355–60.

20. Silberbogen AK, Janke EA, Hebenstreit C. A closer look at pain and hepatitis C: preliminary data from a veteran population. J Rehabil Res Dev 2007; 44:231–44.

21. Whitehead AJ, Dobscha SK, Morasco BJ, Ruimy S, Bussell C, Hauser P. Pain, substance use disorders and opioid analgesic prescription patterns in veterans with hepatitis C. J Pain Symptom Manage 2008.

22. Thompson ME, Barkhuizen A. Fibromyalgia, hepatitis C infection, and the cytokine connection. Curr Pain Headache Rep 2003; 7:342–7.

23. Narvaez J, Nolla JM, Valverde-Garcia J. Lack of association of fibromyalgia with hepatitis C virus infection. J Rheumatol 2005; 32:1118–21.

24. Caldeiro RM, Malte CA, Calsyn DA, Baer JS, Nichol P, Kivlahan DR, Saxon AJ. The association of persistent pain with out-patient addiction treatment outcomes and service utilization. Addiction 2008; 103:1996–2005.

25. Miller I, Bishop S, Norman W, Maddever H. The modified Hamilton rating scale for depression: reliability and validity. Psych Res 1985; 14:131–42.

26. Gerbert B, Bronstone A, McPhee S, Pantilat S, Allerton M. Development and testing of an HIV-risk screening instrument for use in health care settings. Am J Prev Med 1998; 15:103–13.

27. Melzack R. The short-form McGill Pain Questionnaire. Pain 1987; 30:191–7.

28. Cleeland, CS, Ryan, KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singapore 1994; 23:129–38.

29. Schmidt NB, Richey JA, Fitzpatrick KK. Discomfort intolerance: development of a construct and measure relevant to panic disorder. J Anxiety Disord 2006; 20:263–80.

30. Beck AT, Steer RA, Brown GK. Manual for the BDI-II. San Antonio, TX: The Psychological Corporation, 1996.

31. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. Arch Intern Med 2003; 163:2433–45.

32. Libman H, Saitz R, Nunes D, Cheng DM, Richardson JM, Vidaver J, Alperen JK, Samet JH. Hepatitis C infection is associated with depressive symptoms in HIV-infected adults with alcohol problems. Am J Gastroenterol 2006; 101:1804–10.

33. Milligan ED, Ledeboer A, Sloane EM, Maier SF, Busha DA, Watkins LR. Glially driven enhancement of pain and its control by anti-inflammatory cytokines. In: DeLeo JA, Sorkin LS, Watkins LR, eds. Immune and glial regulation of pain. Seattle: IASP Press, 2007.

34. Watkins LR, Maier SF. Immune regulation of central nervous system functions: from sickness responses to pathological pain. J Intern Med 2005; 257:139–55.

35. Knobler H, Zhornicky T, Sandler A, Haran N, Ashur Y, Schattner A. Tumor necrosis factor-alpha-induced insulin resistance may mediate the hepatitis C virus-diabetes association. Am J Gastroenterol 2003; 98:2751–6.

36. Loftis JM, Huckans M, Ruimy S, Hinrichs DJ, Hauser P. Depressive symptoms in patients with chronic hepatitis C are correlated with elevated plasma levels of interleukin-1beta and tumor necrosis factor-alpha. Neurosci Lett 2008; 430:264–8.

37. Malaguarnera M, Di Fazio I, Romeo MA, Restuccia S, Laurino A, Trovato BA. Elevation of interleukin 6 levels in patients with chronic hepatitis due to hepatitis C virus. J Gastroenterol 1997; 32:211–5.

38. Migita K, Abiru S, Maeda Y, Daikoku M, Ohata K, Nakamura M, Komori A, Yano K, Yatsuhashi H, Eguchi K, Ishibashi H. Serum levels of interleukin-6 and its soluble receptors in patients with hepatitis C virus infection. Hum Immunol 2006; 67:27–32.

39. Nelson DR, Lim HL, Marousis CG, Fang JW, Davis GL, Shen L, Urdea MS, Kolberg JA, Lau JY. Activation of tumor necrosis factor-alpha system in chronic hepatitis C virus infection. Dig Dis Sci 1997; 42:2487–94.

40. Oyanagi Y, Takahashi T, Matsui S, Takahashi S, Boku S, Takahashi K, Furukawa K, Arai F, Asakura H. Enhanced expression of interleukin-6 in chronic hepatitis C. Liver 1999; 19:464–72.

41. Huang YS, Hwang SJ, Chan CY, Wu JC, Chao Y, Chang FY, Lee SD. Serum levels of cytokines in hepatitis C-related liver disease: a longitudinal study. Zhonghua Yi Xue Za Zhi (Taipei) 1999; 62:327–33.

42. Stein MD, Maksad J, Clarke J. Hepatitis C disease among injection drug users: knowledge, perceived risk and willingness to receive treatment. Drug Alcohol Depend 2001; 61:211–5.

PAIN® 152 (2011) 2640–2644

www.elsevier.com/locate/pain

Escitalopram is associated with reductions in pain severity and pain interference in opioid dependent patients with depressive symptoms

Judith I. Tsui [a,*], Debra S. Herman [b,c], Malyna Kettavong [b,c], Bradley J. Anderson [b,c], Michael D. Stein [b,c]

[a] Boston University School of Medicine/Boston Medical Center, 801 Massachusetts Ave, Boston MA 02118, USA
[b] Butler Hospital, 345 Blackstone Blvd, Providence, RI 02906, USA
[c] Warren Alpert School of Medicine at Brown University, Waterman Street, Providence, RI, 02912, USA

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

ARTICLE INFO

Article history:
Received 31 March 2011
Received in revised form 14 July 2011
Accepted 8 August 2011

Keywords:
Anti-depressant
Escitalopram
Opioid dependence
Pain

ABSTRACT

Pain is common among opioid-dependent patients, yet pharmacologic strategies are limited. The aim of this study was to explore whether escitalopram, a selective serotonin reuptake inhibitor, was associated with reductions in pain. The study used longitudinal data from a randomized, controlled trial that evaluated the effects of escitalopram on treatment retention in patients with depressive symptoms who were initiating buprenorphine/naloxone for treatment of opioid dependence. Participants were randomized to receive escitalopram 10 mg or placebo daily. Changes in pain severity, pain interference, and depression were assessed at 1-, 2-, and 3-month visits with the visual analog scale, Brief Pain Inventory, and the Beck Depression Inventory II, respectively. Fixed-effects estimators for panel regression models were used to assess the effects of intervention on changes in outcomes over time. Additional models were estimated to explore whether the intervention effect was mediated by within-person changes in depression. In this sample of 147 adults, we found that participants randomized to escitalopram had significantly larger reductions on both pain severity ($b = -14.34$, $t = -2.66$, $P < .01$) and pain interference ($b = -1.20$, $t = -2.23$, $P < .05$) between baseline and follow-up. After adjusting for within-subject changes in depression, the estimated effects of escitalopram on pain severity and pain interference were virtually identical to the unadjusted effects. This study of opioid-dependent patients with depressive symptoms found that treatment with escitalopram was associated with clinically meaningful reductions in pain severity and pain interference during the first 3 months of therapy.

1. Introduction

Pain is common in opioid-dependent patients. Among methadone-treated patients, estimates of chronic pain prevalence range 37–61% [3,14,21]. Several hypothetical pathways may lead to the coexistence of pain and opioid dependence, including depression [2], comorbidities such as HIV [10] or hepatitis C virus [24], and opioid-induced hyperalgesia [8]. Management of pain in opioid-dependent patients is a clinical challenge given concerns for opioid abuse and misuse among individuals with prior substance use disorders [20]. Yet unresolved pain may be a risk factor for relapse among patients whose pain is not fully treated [15]. Some small studies have suggested that buprenorphine/naloxone may be associated with improved pain in opioid-dependent patients with chronic pain [5,16]. However, a study of opioid-dependent patients who were treated with methadone did not find overall changes in pain level at 1 year [13]. Alternative, nonopioid pharmacologic therapies are needed to address pain in opioid-dependent populations.

Antidepressants may constitute an appealing option for treating pain in opioid-dependent patients because of the frequent coexistence of depression in this population [22]. Systematic reviews and clinical guidelines support the use of antidepressants as pharmacotherapy for chronic pain conditions such as low back pain [7], fibromyalgia [27], and neuropathy [23], with the bulk of research to date being focused on use of tricyclic antidepressants and serotonin-norepinephrine reuptake inhibitors. The use of selective serotonin reuptake inhibitors (SSRIs) for chronic pain conditions has been less well studied. Escitalopram belongs to a class of newer SSRIs. It is the S-enantiomer of the SSRI citalopram that has been shown to be responsible for the drug's pharmacologic effect. Two small studies have reported escitalopram to be effective in treating pain in the setting of polyneuropathy [19] and low back pain [17]. No studies have evaluated the effects of escitalopram on pain in opioid-dependent populations.

* Corresponding author. Tel.: +1 617 414 6912; fax: +1 617 414 4676.
E-mail address: Judith.tsui@bmc.org (J.I. Tsui).

0304-3959/$36.00
doi:10.1016/j.pain.2011.08.011

We recently completed a clinical trial of escitalopram for the treatment of depressive symptoms to reduce treatment dropout in opioid-dependent persons initiating buprenorphine [26]. This secondary analysis was undertaken to examine the effects of escitalopram on pain severity and pain interference in that sample.

2. Methods

2.1. Study sample and design

This study used longitudinal data from a randomized, controlled trial that evaluated whether treatment with escitalopram increased treatment retention among opioid-dependent patients with depressive symptoms who were initiating buprenorphine/naloxone [26]. Participants were recruited through community advertising, physician referrals, and word of mouth. Study inclusion criteria included: age 18–65, DSM-IV diagnosis of opioid dependence, score on the Modified Hamilton Depression Revised Scale of >14 [18], absence of significant suicidal ideation, willingness and ability to complete a 3-month treatment with buprenorphine, no history of severe mental illness (bipolar disorder, schizophrenia, schizoaffective, or paranoid disorder), no currently prescribed medications for depression (participants were not excluded if they were receiving a tricyclic antidepressant for pain), and ability to complete the study assessment in English. The study was approved by the Rhode Island Hospital and Butler Hospital institutional review boards.

Between November 2006 and May 2009, a total of 932 individuals were screened by telephone, and of those, 394 callers seemed eligible for the study and were invited for an in-person screening visit. Of the 226 who attended this visit, 147 fully met the study criteria and agreed to enroll onto the parent study. Of these 147, a total of 72 were randomized to the intervention with escitalopram and 75 were randomized to placebo. In the intervention arm, 48 completed the week 12 study visit, and in the placebo arm, 42 completed the week 12 study visit. The reason for loss to follow-up was dropout from buprenorphine treatment.

Participants who enrolled onto the study completed a baseline interview that included questions on pain, and they were randomized to receive the study medication escitalopram 10 mg or placebo daily. Double-blinding regarding the medication group was maintained throughout the study period for all research staff; the off-site compounding pharmacist kept the key to the blind. Study medication and placebo were provided in identical capsule form. Approximately 5 days after beginning the study medication, participants returned to the research office for buprenorphine (buprenorphine/naloxone) induction. Dose adjustments were based on previous opioid use, craving, and reported symptoms. In general, buprenorphine doses ranging 12–24 mg/day were required for stabilization. At follow-up interviews at 1, 2, and 3 months after enrollment, participants were again asked about their pain.

2.2. Measures

The primary outcomes measured were pain severity and pain interference in the past week. Pain severity was measured with the visual analog scale (VAS) [6]. Participants were asked to rate their pain by placing a mark on a 100-mm nonhatched line that was marked as "No pain" at one end and "Pain as bad as you can imagine" on the other end. Pain interference was assessed with the mean of the 7-item subscale from the Brief Pain Inventory Short Form (BPI) [9]. This subscale measures pain interference in different domains such as sleep, work, and relationships, rating each item from 0 (pain does not interfere) to 10 (pain completely interferes). Other measures included age, sex, race/ethnicity, educational status, primary illicit opioid used, withdrawal pain (number of days experienced in the past week), and depressive symptoms. Depressive symptoms were measured with the Beck Depression Inventory II (BDI II) [4].

2.3. Analytical methods

Descriptive statistics are presented to summarize the characteristics of the study cohort; *t* tests and Pearson chi-square tests are presented to compare intervention arms on a range of background characteristics and indicators of study attrition. We used graphical methods to describe the pattern of pain severity and pain interference observed over time. We used the fixed-effects estimator [25] for panel regression models to statistically analyze change in pain over time and to assess the effects of intervention on change in these outcomes over time. The fixed-effects estimator uses only within-subject variability and effectively controls between subject heterogeneity on all time-invariant characteristics [1]. A potential limitation is that the effect of time constant between subject characteristics (eg, sex, ethnicity) on outcomes cannot be estimated. However, the fixed-effects estimator effectively controls for all unmeasured between subject differences that do not change over time [1]. All confidence interval estimates and tests of significance were based on the robust standard error estimators as implemented in Stata 10.1 [25]. The effect of time-invariant characteristics (treatment assignment) on change in the outcome over time can be estimated as the predictor by time interaction.

Our analysis proceeded in stages. We first estimated unconditional growth models to evaluate the change process. We used the likelihood-ratio difference in the chi-square test to compare the linear growth model to an unconstrained time model estimating separate parameters for each follow-up using 3 dummy variables. Examination of results suggested a more parsimonious parameterization in which time was represented by a single dummy variable coded 0 if baseline and 1 if follow-up. Conditional growth models that included the treatment by time interaction were specified to estimate the effects of intervention on change in pain severity and pain interference. Additional fixed-effects regression models were estimated to test the hypothesis that the effect of escitalopram on pain was mediated by within-person changes in depression.

3. Results

In our sample of 147 adults, participants averaged a mean ± standard deviation (SD) age of 37.5 ± 9.9 years, and most were men (76%) (Table 1). Most participants (80.1%) were non-Hispanic white, 4.9% were African American, 9.6% were Hispanic, and 5.5% were of other racial or ethnic origins. Ninety-three (63.7%) said heroin was their opiate of choice. The mean BDI II score at baseline was 29.4 ± 9.7. Most participants (85%) would be classified as having moderate to severe depression [4]. By means of the Structured Clinical Interview for DSM Disorders [12], 51% met the criteria for current major depression and 4% for dysthymia. Most (122, 83%) reported some pain in the past week, and of those with pain, 48% had pain that was chronic ($\geqslant$6 months). Among participants who reported pain, the most common sites were back pain (66%), joint pain (48%), and muscle pain (38%). Over 90% of participants were located for at least 1 follow-up, and there was no evidence of significant between-group differences with respect to attrition at any time point (Table 1). Intervention groups also did not differ significantly with respect to other background characteristics and pain. There was no evidence of systematic differences with respect to study attrition.

In the full cohort, mean VAS and BPI scores declined from baseline to 1 month and stayed relatively constant at months 2 and 3 (Fig. 1). Statistical comparisons of mean VAS and BPI scores across

Table 1
Baseline characteristics by intervention (n = 147).

Characteristic	Placebo (n = 75)	Escitalopram (n = 72)	P
Age, y, mean ± SD	36.8 ± 9.8	38.3 ± 9.6	.39
Education, y, mean ± SD	12.5 ± 1.8	11.9 ± 1.6	.06
Ethnicity, n (%)			
White	60 (81.1%)	57 (79.2%)	
African American	3 (4.0%)	4 (5.6%)	
Hispanic	7 (9.5%)	7 (9.7%)	
Other	4 (5.4%)	4 (5.6%)	.98
Male	57 (76.0%)	55 (76.4%)	.90
Heroin user, n (%)	44 (59.5%)	49 (68.1%)	.28
Baseline pain type, n (%)			.125
None	17 (22.7%)	8 (11.1%)	
Pain, not chronic	22 (29.3%)	29 (40.3%)	
Chronic pain	36 (48.0%)	35 (48.6%)	
In the past week, no. of days with pain due to withdrawal, mean ± SD	3.3 ± 2.7	2.9 ± 2.6	.35
BDI score, mean ± SD	28.6 ± 9.8	28.3 ± 9.6	.83
VAS, mean ± SD	44.6 ± 31.8	54.1 ± 31.8	.08
BPI, mean ± SD	3.6 ± 3.0	4.5 ± 2.8	.06
Observed at:			
1 month, *n* (%)	64 (85.3%)	64 (88.9%)	.52
2 months, *n* (%)	51 (68.0%)	53 (73.6%)	.46
3 months, *n* (%)	48 (64.0%)	42 (58.3%)	.48
⩾1 Follow-up, *n* (%)	69 (92.0%)	69 (95.8%)	.33

the 1-, 2-, and 3-month follow-up visits were not significant (ie, $P > .10$) for all pairwise comparisons. Because differences in mean VAS and BPI scores across follow-up assessments were substantively small and not statistically significant, we also estimated a more parsimonious model constraining the outcome means to be equal across all follow-up assessments. Likelihood-ratio (LR) chi-square difference tests indicated that the more complex unconstrained time model did not fit the data significantly better than the simpler parameterization of time for either the VAS ($LR^2 = 1.41$, df = 2, $P = 0.50$) or BPI ($LR^2 = 3.54$, df = 2, $P = 0.17$); therefore, we used this parameterization for the subsequent analysis.

Table 2, Model 1, gives the results of fixed-effects regression estimating the effect of intervention on VAS and BPI at follow-up. The coefficient for time gives the estimated change in outcomes between baseline and follow-up in the placebo group among whom mean VAS and BPI scores decreased by 16.8 ($t = -4.50$, $P < .01$) and 1.15 ($t = -3.69$, $P < .01$) points, respectively, between baseline and follow-up. The "treatment by time" coefficient represents the effect of escitalopram relative to placebo. Compared to those receiving placebo, participants who were randomized to escitalopram had 14.34 ($t = -2.66$, $P < .01$) and 1.20 ($t = -2.23$, $P < .05$) larger mean reductions in VAS and BPI, respectively. Fig. 2 provides the expected follow-up mean VAS and BPI scores by treatment arm, as estimated by the fixed-effects model. Mean VAS scores a follow-up were estimated to be 32.1 (95% confidence interval [CI] 26.4–37.8) and 17.7 (95% CI 12.2–23.2) among those randomized to placebo and escitalopram, respectively. Predicted mean BPI scores were 2.53 (95% CI 1.96–3.10) for those receiving placebo and 1.33 (95% CI 0.79–1.87) in the escitalopram arm.

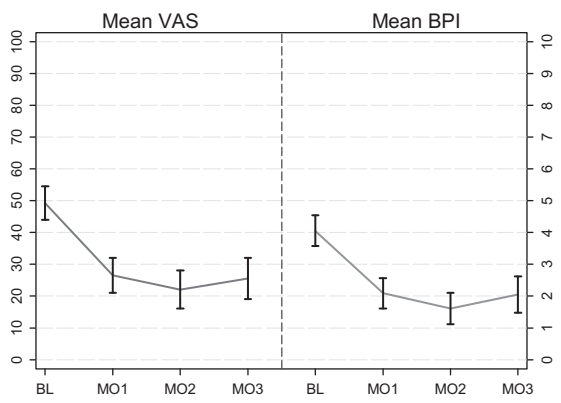


Fig. 1. Change in pain severity (VAS) and pain interference (BPI) from baseline to 3 months. Error bars denote 95% confidence interval estimates of the mean.

Because changes in depression represent a plausible mechanism through which the effect of escitalopram on pain may be mediated, we also estimated a model that included BDI II as a time-varying covariate (Table 2, Model 2). If changes in depression mediated the effect of escitalopram on pain, we would have expected that the effect of escitalopram on pain would be attenuated. We found no support for the mediation hypothesis. After adjusting for within-subject changes in depression, the estimated effect of escitalopram on pain severity ($b = -14.37$, $t = -2.69$, $P < .01$) and pain interference ($b = -1.20$, $t = -2.30$, $P < .05$) were virtually identical to the unadjusted effects reported in Model 1. However, within-subject change in BDI II was associated significantly with within-subject change in VAS pain severity ($t = 2.52$, $P < .05$) and BPI pain interference ($t = 4.25$, $P < .01$). A within-subject increase of 1 point on the BDI II was associated with a .55 point increase in mean VAS and a .09 point increase in mean BPI scores across time.

4. Discussion

This study of opioid-dependent patients with depressive symptoms found that treatment with escitalopram resulted in significantly decreased pain severity and interference over time. Adjusting for within-subject changes in depression scores did not affect the effects of escitalopram, suggesting the analgesic properties of escitalopram were independent of its antidepressant effects.

To our knowledge, this is the first study that demonstrates an association between the antidepressant escitalopram and improved general pain. A small randomized, controlled crossover

Table 2
Fixed-effects regression models estimating the effect of escitalopram on pain severity and pain interference, without (model 1) and with (model 2) adjustment for depression.

Characteristic	Pain severity		Pain interference	
	Model 1 β (SE)	Model 2 β (SE)	Model 1 β (SE)	Model 2 β (SE)
Time[a]	−16.82** (3.74)	−8.37 (4.50)	−1.48** (0.40)	−0.16 (0.51)
Treatment by time[b]	−14.34** (5.39)	−14.37** (5.34)	−1.20* (0.54)	−1.20* (0.52)
BDI II (time varying)	NA	0.55* (0.22)	NA	0.09** (0.02)
Intercept	48.89	33.47	4.01	1.60

[a] Coefficient gives the expected change in pain between baseline and follow-up visits for participants randomized to placebo.
[b] Relative to controls, this coefficient gives the expected mean change between baseline and follow-up visits for participants randomized to escitalopram.
* $P < .05$.
** $P < .01$.

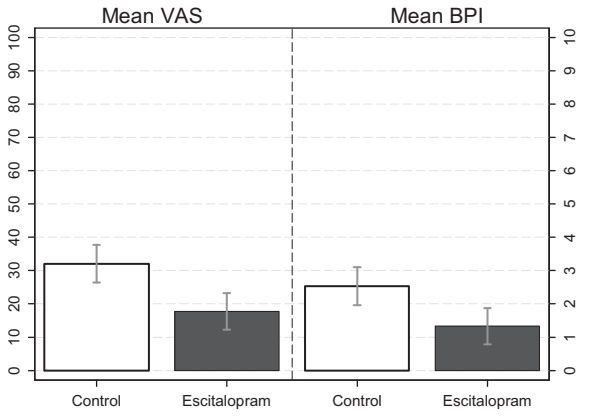


Fig. 2. Predicted follow-up mean VAS and BPI scores by intervention group. Predicted values were estimated by the fixed-effects estimator (Table 2, Model 1); error bars denote 95% confidence interval estimates of the predicted mean.

study of patients with painful peripheral neuropathy found that escitalopram was associated with significantly lower pain ratings, and that the treatment effect was concentrated among patients with hyperalgesia [19]. Another randomized, controlled trial compared treatment with escitalopram to duloxetine (which is approved by the US Food and Drug Administration for treatment of diabetic neuropathy, fibromyalgia, and chronic musculoskeletal pain) in patients with chronic low back pain, and found it to be equally efficacious for reducing pain [17]. Although our results demonstrated statistical significance, it is also important to assess clinical significance. In our study, the intervention escitalopram was associated with a 14-point decrease beyond what was observed with placebo, which reflects a 29% improvement from the baseline mean of 49. Recent consensus guidelines suggest that 10–20% and $\geqslant$30% decreases in numeric pain intensity represent minimally and moderately important changes in pain, respectively [11]. Therefore, our results suggest that escitalopram is associated with improvement in pain severity that exceeds minimal clinical importance. The same guidelines suggest that a change of 1 point of the BPI interference scale should be considered the threshold for clinically important change in pain interference. By means of this criterion, our results, which show a 1.2-point change associated with escitalopram, also suggest a clinically significant effect on pain interference.

The effects of escitalopram on pain in this study did not seem to depend on the drug's antidepressant effect. This was observed by adjusting for within-subject changes in depression in the second model. Depression scores did not differ significantly between the placebo and escitalopram over the course of the study (possibly related to the relatively low dose of escitalopram) [26]. This underscores the conclusion that the escitalopram's effect on reducing pain in this study was independent of the drug's antidepressant effect. However, the study also demonstrated that there were improvements in depressive symptoms over time in both placebo and intervention arms that were associated with reductions in pain. This reinforces the close relationship between pain and depression [2], which is relevant because depressive symptoms are highly prevalent in this population [22]. It is also of interest to note that pain decreased in the placebo group during the study period. This could be due to the placebo effect, the effects of buprenorphine, the resolution of withdrawal pain, or, most likely, some combination of all factors. Nonetheless, additional pain relief was observed in the escitalopram arm.

This study has numerous limitations and strengths. The study is based on secondary analysis of data; however, it makes use of a randomized, controlled study design with a blinded intervention. At baseline, the intervention group had slightly higher (though nonsignificant) mean VAS and BPI severity scores, but this should not influence the study's main findings, which focus on the effects of the intervention on change in VAS and BPI over time. The study had a relatively short follow-up time of 3 months. It is unknown whether reductions in pain associated with escitalopram are sustained beyond this period. The dose of escitalopram used in this study was relatively low (10 mg). It is possible that larger reductions in pain might have occurred with use of a higher dose. The study is focused on a relatively specific patient population, namely opioid-dependent patients with depressive symptoms initiating buprenorphine, which may limit the generalizability of the findings. However, there are studies that support escitalopram's effectiveness treating pain in non-opioid-dependent populations [19,17]. Furthermore, opioid-dependent patients are disproportionately affected by pain (yet are rarely included in clinical trials), and alternatives to narcotic medications are greatly needed for this population.

In summary, this study of opioid-dependent patients with depressive symptoms found that treatment with escitalopram was associated with a reduction in pain severity and pain interference during the first 3 months of buprenorphine therapy. Treatment with escitalopram was associated with a nearly 30% reduction in pain severity after 1 month compared to control, and its analgesic effect seemed to be independent from any antidepressant effect. More research is needed on the use of nonnarcotic medications such as SSRIs to treat pain in opioid-dependent populations.

Conflict of interest statement

The authors declare no conflicts of interest related to this study.

Acknowledgments

This study was funded by the National Institute on Drug Abuse DA 022207, Clinical Trial NCT 00475878. MDS is a recipient of a

NIDA Mid-Career Award DA 000512. JIT is a recipient of a NIDA Career Development Award DA 027367.

References

[1] Allison PD. Fixed effects regression methods for longitudinal data using SAS. Cary, NC: SAS Institute; 2005.

[2] Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. Arch Intern Med 2003;163:2433–45.

[3] Barry DT, Beitel M, Garnet B, Joshi D, Rosenblum A, Schottenfeld RS. Relations among psychopathology, substance use, and physical pain experiences in methadone-maintained patients. J Clin Psychiatry 2009;70:1213–8.

[4] Beck AT, Steer RA, Brown GK. Manual for the BDI-II. San Antonio, TX: The Psychological Corporation; 1996.

[5] Blondell RD, Ashrafioun L, Dambra CM, Foschio EM, Zielinski AL, Salcedo DM. A clinical trial comparing tapering doses of buprenorphine with steady doses for chronic pain and co-existent opioid addiction. J Addict Med 2010;4:140–6.

[6] Breivik H, Borchgrevink PC, Allen SM, Rosseland LA, Romundstad L, Hals EK, Kvarstein G, Stubhaug A. Assessment of pain. Br J Anaesth 2008;101:17–24.

[7] Chou R, Huffman LH. Medications for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. Ann Intern Med 2007;147:505–14.

[8] Chu LF, Angst MS, Clark D. Opioid-induced hyperalgesia in humans: molecular mechanisms and clinical considerations. Clin J Pain 2008;24:479–96.

[9] Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singapore 1994;23:129–38.

[10] Dobalian A, Tsao JC, Duncan RP. Pain and the use of outpatient services among persons with HIV: results from a nationally representative survey. Med Care 2004;42:129–38.

[11] Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, Haythornthwaite JA, Jensen MP, Kerns RD, Ader DN, Brandenburg N, Burke LB, Cella D, Chandler J, Cowan P, Dimitrova R, Dionne R, Hertz S, Jadad AR, Katz NP, Kehlet H, Kramer LD, Manning DC, McCormick C, McDermott MP, McQuay HJ, Patel S, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Revicki DA, Rothman M, Schmader KE, Stacey BR, Stauffer JW, von Stein T, White RE, Witter J, Zavisic S. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105–21.

[12] First M, Spitzer R, Williams J, Gibbons M. Structured clinical interview for DSM-IV—patient version. New York: New York State Psychiatric Institute; 1995.

[13] Ilgen MA, Trafton JA, Humphreys K. Response to methadone maintenance treatment of opiate dependent patients with and without significant pain. Drug Alcohol Depend 2006;82:187–93.

[14] Jamison RN, Kauffman J, Katz NP. Characteristics of methadone maintenance patients with chronic pain. J Pain Symptom Manage 2000;19:53–62.

[15] Larson MJ, Paasche-Orlow M, Cheng DM, Lloyd-Travaglini C, Saitz R, Samet JH. Persistent pain is associated with substance use after detoxification: a prospective cohort analysis. Addiction 2007;102:752–60.

[16] Malinoff HL, Barkin RL, Wilson G. Sublingual buprenorphine is effective in the treatment of chronic pain syndrome. Am J Ther 2005;12:379–84.

[17] Mazza M, Mazza O, Pazzaglia C, Padua L, Mazza S. Escitalopram 20 mg versus duloxetine 60 mg for the treatment of chronic low back pain. Expert Opin Pharmacother 2010;11:1049–52.

[18] Miller I, Bishop S, Norman W, Maddever H. The modified Hamilton rating scale for depression: reliability and validity. Psych Res 1985;14:131–42.

[19] Otto M, Bach FW, Jensen TS, Brosen K, Sindrup SH. Escitalopram in painful polyneuropathy: a randomized, placebo-controlled, cross-over trial. Pain 2008;139:275–83.

[20] Passik SD. Issues in long-term opioid therapy: unmet needs, risks, and solutions. Mayo Clin Proc 2009;84:593–601.

[21] Rosenblum A, Joseph H, Fong C, Kipnis S, Cleland C, Portenoy RK. Prevalence and characteristics of chronic pain among chemically dependent patients in methadone maintenance and residential treatment facilities. JAMA 2003;289:2370–8.

[22] Rounsaville BJ, Weissman MM, Crits-Christoph K, Wilber C, Kleber H. Diagnosis and symptoms of depression in opiate addicts. Course and relationship to treatment outcome. Arch Gen Psychiatry 1982;39:151–6.

[23] Saarto T, Wiffen PJ. Antidepressants for neuropathic pain: a Cochrane review. J Neurol Neurosurg Psychiatry 2010;81:1372–3.

[24] Silberbogen AK, Janke EA, Hebenstreit C. A closer look at pain and hepatitis C: preliminary data from a veteran population. J Rehabil Res Dev 2007;44:231–44.

[25] StataCorp. Stata statistical software: release 10.1. College Station, TX: StataCorp, 2010.

[26] Stein MD, Herman DS, Kettavong M, Cioe PA, Friedmann PD, Tellioglu T, Anderson BJ. Antidepressant treatment does not improve buprenorphine retention among opioid-dependent persons. J Subst Abuse Treat 2010;39:157–66.

[27] Uceyler N, Hauser W, Sommer C. A systematic review on the effectiveness of treatment with antidepressants in fibromyalgia syndrome. Arthritis Rheum 2008;59:1279–98.

Comparison of Diabetes Control Among Haitians, African Americans, and Non-Hispanic Whites in an Urban Safety-Net Hospital

VARSHA G. VIMALANANDA, MD, MPH[1]
JAMES L. ROSENZWEIG, MD[1]
HOWARD J. CABRAL, PHD, MPH[2]
MICHELE M. DAVID, MD, MPH, MBA[3]
KAREN E. LASSER, MD, MPH[4,5]

OBJECTIVE — To compare diabetes care and outcomes among Haitians, African Americans, and non-Hispanic whites.

RESEARCH DESIGN AND METHODS — We analyzed data from 715 Haitian, 1,472 African American, and 466 non-Hispanic white adults with diabetes using χ^2 testing and multiple logistic regression.

RESULTS — Haitians had a higher mean A1C than African Americans (8.2 ± 1.9 vs. 7.7 ± 2.0%) and non-Hispanic whites (7.5 ± 1.7%) (both $P < 0.0001$). There was no difference in completion of process measures. Haitians were more likely than non-Hispanic whites to have elevated LDL cholesterol or blood pressure. Macrovascular complications were fewer among Haitians than African Americans (adjusted odds ratio 0.35 [95% CI 0.23–0.55]), as were microvascular complications (0.56 [0.41–0.76]). Haitians also had fewer macrovascular (0.32 [0.20–0.50]) and microvascular (0.55 [0.39–0.79]) complications than non-Hispanic whites.

CONCLUSIONS — Haitians have worse glycemic control than African Americans or non-Hispanic whites. Future research and interventions to improve diabetes care should target Haitians as a distinct racial/ethnic group.

Diabetes Care 34:58–60, 2011

There are 531,000 black individuals of Haitian ancestry living in the U.S. (1). We identified no studies of diabetes care or outcomes in this population. Thus, it is unclear whether Haitians, like African Americans, have a higher mean A1C (2), receive less recommended testing (3), or have higher rates of retinopathy (4), nephropathy (5), or lower extremity amputations (6) than whites. We analyzed data from primary care clinics in the largest safety-net hospital in Massachusetts in order to compare diabetes care and outcomes among Haitians, African Americans, and non-Hispanic whites.

RESEARCH DESIGN AND METHODS

— We conducted an observational study of subjects with diabetes who received primary care at Boston Medical Center, an urban safety-net hospital with academic primary care practices. The Boston University Medical Center institutional review board approved the study protocol. We included individuals who had at least one primary care visit yearly between 1 August 2007 and 1 August 2009, were ≥20 years old, carried a diagnosis of diabetes by ICD-9-CM billing code 250.XX or by presence on the medical record problem list, and who self-identified as either Haitian, African American, or non-Hispanic white.

The percentage of patients with poor glycemic control (A1C >9%) (7) was our primary outcome measure. Process measures included yearly testing of A1C, LDL cholesterol, and urine microalbumin in patients without nephropathy (8). We ascertained diabetic complications by ICD-9-CM codes or presence on the problem list, although we defined nephropathy as an estimated glomerular filtration rate <60 ml/min, urinary albumin-to-creatinine ratio ≥30 mg/g, or a history of kidney transplant or dialysis.

We used SAS statistical software (version 9.1; SAS Institute, Cary, NC), performing cross-tabulations with χ^2 tests where appropriate and multiple logistic regression to analyze race/ethnicity as a predictor of outcomes. Each regression model included age, sex, language, insurance type, number of primary care visits over 2 years, and having at least one visit to an endocrinologist over 2 years. In models of complication risk, we also controlled for BMI, hypertension diagnosis, and having ever smoked. We assessed each model for interactions between race/ethnicity and either sex or higher health care utilization. A similar analysis was performed to compare English- and non-English–speaking Haitians.

RESULTS

— We identified 2,653 subjects, including 715 Haitians, 1,472 African Americans, and 466 non-Hispanic whites. Thirty-two percent of Haitians were English-speaking. Haitians were of similar mean age to African Americans (58.8 ± 12.0 vs. 57.8 ± 12.5 years) and non-Hispanic whites (59.8 ± 11.8 years), but had a lower mean BMI compared with both African Americans (30.8 ± 6.0 vs. 33.8 ± 8.0 kg/m^2, $P < 0.05$) and non-Hispanic whites (33.4 ± 8.0 kg/m^2, $P < 0.05$). A history of smoking was signifi-

From the [1]Department of Medicine, Section of Endocrinology, Diabetes and Nutrition, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts; the [2]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts; the [3]Department of Medicine, Women's Health Research Unit, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts; the [4]Department of Medicine, Section of General Internal Medicine, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts; and the [5]Department of Social and Behavioral Science, Boston University School of Public Health, Boston, Massachusetts.
Corresponding author: Varsha G. Vimalananda, varsha.vimalananda@bmc.org.
Received 19 July 2010 and accepted 11 October 2010. Published ahead of print at http://care.diabetesjournals.org on 26 October 2010. DOI: 10.2337/dc10-1387.

Table 1—*Intermediate outcomes and diabetic complications by race/ethnicity*

	Haitians	African Americans	*P**	aOR (95% CI)†‡	Non-Hispanic whites	*P*§	aOR (95% CI)‖‡
n	715	1,472			466		
Intermediate outcome measures							
A1C >9% (%)	24	18	0.003	1.43 (1.04–2.00)	15	0.0002	1.67 (1.11–2.50)
LDL ≥100 mg/dl (%)	29	30	0.61	1.08 (0.80–1.45)	19	<0.0001	1.85 (1.28–2.63)
BP ≥140/80 mmHg (%)¶	51	45	0.01	1.07 (0.81–1.43)	33	<0.0001	1.85 (1.30–2.50)
Diabetic complications							
Macrovascular#	20	38	<0.0001	0.35 (0.23–0.55)	42	<0.0001	0.32 (0.20–0.50)
Microvascular**	46	59	<0.0001	0.56 (0.41–0.76)	61	<0.0001	0.55 (0.39–0.79)

**P* value for Haitians versus African Americans. †Adjusted odds ratio for Haitians versus African Americans. ‡Odds ratios adjusted for age, sex, language (English speaking/non-English speaking), and insurance type (Medicaid or Free Care, Medicare, private, and other insurance), number of primary care visits over 2 years, and having at least one endocrinologist visit over 2 years. Models for complications are additionally adjusted for BMI, diagnosis of hypertension, and ever having smoked. §*P* value for Haitians versus non-Hispanic whites. ‖Adjusted odds ratio for Haitians versus non-Hispanic whites. ¶This quality measure selected on the basis of clinical trials, which show a reduction in coronary heart disease events, stroke, and nephropathy with blood pressure <140/80 mmHg (9). #Macrovascular complications include coronary artery disease, congestive heart failure, ischemic stroke, peripheral vascular disease, and lower extremity ulcers. **Microvascular complications include retinopathy, nephropathy, and neuropathy. aOR, adjusted odds ratio; BP, blood pressure.

cantly less common among Haitians compared with African Americans (52 vs. 85%, $P < 0.05$) and non-Hispanic whites (77%, $P < 0.05$). Compared with African Americans, Haitians had lower health care utilization as measured by number of primary care visits over 2 years (9.2 ± 4.7 vs. 9.8 ± 5.9, $P < 0.05$) and the likelihood of having an endocrinologist visit (16 vs. 26%, $P < 0.05$).

The mean A1C was higher among Haitians than among African Americans (8.2 ± 1.9 vs. 7.7 ± 2.0%, $P < 0.0001$) and among non-Hispanic whites (7.5 ± 1.7%, $P < 0.0001$), and the higher risk of poor glycemic control among Haitians persisted after adjustment (Table 1). In the unadjusted analysis, Haitians had a higher risk of poor blood pressure control compared with both groups and a higher risk of poor LDL cholesterol control compared with non-Hispanic whites. After adjustment, these differences persisted in the comparison with non-Hispanic whites only. Rates of process measure completion were comparable across groups. The prevalence of retinopathy was similar across groups, but all other complications were less common among Haitians. Compared with African Americans, Haitians had lower adjusted odds ratios for macrovascular and microvascular complications (0.35 [95% CI 0.23–0.55] and 0.56 [0.41–0.76], respectively). These risks were also lower among Haitians than they were among non-Hispanic whites (0.32 [0.20–0.50] and 0.55 [0.39–0.79]). In the analysis of nephropathy alone, we found that Haitians fared better than both African Americans (0.56 [0.39–0.80]) and non-Hispanic whites (0.47 [0.31–0.70]). Creole- or French-speaking Haitians had better LDL cholesterol control than English-speaking Haitians, but in the adjusted analysis there were no differences in other outcomes. There were no significant interactions between ethnicity and either sex or health care utilization.

CONCLUSIONS — Haitians had similar rates of completed process measures but worse glycemic control compared with both African Americans and non-Hispanic whites in an urban safety-net hospital. The higher mean A1C among Haitians was evident in both the unadjusted and adjusted analyses, as were the worse lipid and blood pressure control among Haitians compared with non-Hispanic whites. Despite these findings, the rates of diagnosed and documented complications were lower in the Haitian group than in either comparison group.

We identified no other studies of diabetes care and outcomes in Haitians which with to compare these findings. Our results suggest that that worse glycemic control among Haitians may not be attributable to a language barrier or lower health care utilization. Patient-level factors, such as consumption of a traditional high-carbohydrate Haitian diet, or provider- and systems-level factors, such as limited cultural competency, may contribute to worse glycemic control among Haitians.

The finding of lower complication rates among Haitians is surprising in light of intermediate outcome measures that are worse than or similar to the comparison groups. One possible explanation is a shorter duration of diabetes among Haitians, but we cannot exclude detection bias or a higher loss to follow-up among this group. That nephropathy was also less common in Haitians is an interesting finding, because this complication was assessed primarily by results of lab testing and presence of serious complications and thus was less subject to underdiagnosis and underdocumentation.

Worse glycemic control is associated with higher risk of hypoglycemia and symptomatic hyperglycemia, and the frequency of complications among Haitians may worsen with increasing acculturation, obesity, and prevalence of diabetes. Future interventions to prevent diabetes-related morbidity and mortality and reduce health disparities should target Haitians and address the unique features of Haitian culture that may affect the course of diabetes care.

Acknowledgments — No potential conflicts of interest relevant to this article were reported.

V.G.V. researched data, contributed to discussion, and wrote the manuscript. J.L.R., H.J.C., M.M.D., and K.E.L. contributed to discussion and reviewed and edited manuscript.

Parts of this study were presented at the 19th Annual Meeting and Clinical Congress of the American Association of Clinical Endocrinologists, Boston, Massachusetts, 21–25 April 2010, and at the New England Regional Meeting of the Society of General Internal Medicine, Providence, Rhode Island, 11 February 2010.

The authors thank Howard Bauchner, MD, of the Boston University Medical Center, Boston, Massachusetts, and Graeme Fincke, MD, of the Center for Health Quality, Outcomes

and Economic Research, Bedford, Massachusetts, for their help in the preparation of this manuscript.

References

1. 2006–2008 American Community Survey 3-year estimates. Table S0201. Selected population profile in the United States: Haitian [Internet], 2008 Washington, DC, U.S. Census Bureau. Available from http://factfinder.census.gov. Accessed 25 May 2010
2. Kirk JK, D'Agostino RB Jr, Bell RA, Passmore LV, Bonds DE, Karter AJ, Narayan KM. Disparities in HbA1c levels between African-American and non-Hispanic white adults with diabetes: a meta-analysis. Diabetes Care 2006;29:2130–2136
3. Schneider EC, Zaslavsky AM, Epstein AM. Racial disparities in the quality of care for enrollees in Medicare managed care. JAMA 2002;287:1288–1294
4. Harris MI. Racial and ethnic differences in health insurance coverage for adults with diabetes. Diabetes Care 1999;22:1679–1682
5. Perneger TV, Brancati FL, Whelton PK, Klag MJ. End-stage renal disease attributable to diabetes mellitus. Ann Intern Med 1994;121:912–918
6. Carter JS, Pugh JA, Monterrosa A. Non-insulin-dependent diabetes mellitus in minorities in the United States. Ann Intern Med 1996;125:221–232
7. National Committee for Quality Assurance (NCQA). *HEDIS 2009: Healthcare Effectiveness Data and Information Set*, vol. 2, *Technical Specifications*. Washington, DC, NCQA, 2008
8. National Quality Forum. *National Voluntary Consensus Standards for Adult Diabetes Care: 2005 Update: A Consensus Report*. Washington, DC, National Quality Forum, 2006
9. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL Jr, Jones DW, Materson BJ, Oparil S, Wright JT Jr, Roccella EJ; National Heart, Lung, and Blood Institute Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure; National High Blood Pressure Education Program Coordinating Committee. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. JAMA 2003;289:2560–2572

JOURNAL OF PALLIATIVE MEDICINE
Volume 14, Number 2, 2011

DOI: 10.1089/jpm.2010.0299

Assessing End-of-Life Preferences for Advanced Dementia in Rural Patients Using an Educational Video: A Randomized Controlled Trial

Angelo E. Volandes, M.D.,[1,2] Laurie Anne Ferguson, D.N.P., APRN,[3] Aretha D. Davis, J.D., MSIV,[2] Nathan C. Hull, MSIV,[4] Michael J. Green, M.D., M.S.,[4] Yuchiao Chang, Ph.D.,[1,2] Kristy Deep, M.D.,[5] and Michael K. Paasche-Orlow, M.D., M.A., M.P.H.[6]

Abstract

Objective: Few studies have evaluated the end-of-life preferences of elderly patients in rural communities and whether preferences are associated with level of health literacy.
Design: Randomized controlled trial of a goals-of-care video decision aid of advanced dementia.
Participants: Elderly subjects (65 years or older) at a primary care clinic in rural Louisiana.
Methods: Half of subjects heard a verbal description of advanced dementia and the goals of care; the other half heard the same verbal description and then viewed the video decision aid. End points were the preferred goal of care in advanced dementia: life-prolonging care (cardiopulmonary resuscitation [CPR], etc.), limited care (hospitalization but not CPR), or comfort care (symptom relief). The principal category for analysis was the difference in proportions of subjects preferring comfort care for each characteristic including randomization group and health literacy level.
Results: Seventy-six subjects were randomized to the verbal ($n = 43$) or video ($n = 33$) arms of the study. Among subjects receiving the verbal description of advanced dementia and the goals of care, 31 (72%) preferred comfort; 5 (12%) chose limited; and 7 (16%) desired life-prolonging. In the video group, 30 (91%) preferred comfort; 3 (9%) chose limited; and none desired life-prolonging ($\chi^2 = 6.3$, $df = 2$, $p = 0.047$). Factors associated with greater likelihood of opting for comfort included greater health literacy (unadjusted odds ratio [OR] 12.1; 95% confidence interval [CI], 2.4–62.6) and randomization to the video (unadjusted OR 3.9; 95% CI, 1.0–15.1).
Conclusion: Rural subjects with higher health literacy were more likely to want comfort care compared to those with lower levels of health literacy. Furthermore, subjects who viewed a video decision aid were more likely to opt for comfort compared to those who solely listened to a verbal description. These findings suggest that video can help elicit preferences and that interventions to empower such patients need to be designed in a manner that is sensitive to health literacy.

Introduction

Despite the fact that rural areas have a larger proportion of seniors over the age of 65 than do urban communities,[1] few studies have explored the end-of-life preferences of rural elderly patients.[2] Instead, the focus of research in end-of-life decision-making has centered around urban populations.[2,3] Nevertheless, given important differences between urban and rural populations, medical practitioners should not assume that taking care of rural patients at the end of life is no different from caring for urban patients at the end of life.[4] With the aging of rural populations, assessing the end-of-life preferences of rural elderly patients is critical for planning and providing palliative care and hospice services for diseases that

[1]General Medicine Unit, Massachusetts General Hospital, Boston, Massachusetts.
[2]Harvard Medical School, Boston, Massachusetts.
[3]School of Nursing, University of Texas Health Science Center, Houston, Texas.
[4]Penn State College of Medicine, Hershey, Pennsylvania.
[5]Department of Medicine, University of Kentucky, Lexington, Kentucky.
[6]Department of Medicine, Boston University School of Medicine, Boston, Massachusetts.
Accepted September 9, 2010.

this underserved population is likely to confront. Furthermore, studying differences in rural–urban end-of-life care is vital for making palliative care services available to all patients regardless of geography.[5,6]

Variations in the use of end-of-life care services between the elderly in rural and urban settings are known to exist,[7,8] with rural populations tending to use fewer medical interventions at the end of life than those in urban settings.[9–13] Such variations may be attributed to a variety of factors including different access to medical services,[3,14–16] expectations of patients,[3] race,[17–25] and rural–urban cultural differences.[2,3,9] The role of culture is increasingly recognized as a vital interpretive lens through which end-of-life decision-making must be understood.[26–28] While much has been written about the impact of cultural differences on end-of-life preferences, there have been few studies focusing on the impact of rural–urban differences on this outcome.

Our own work studying end-of-life preferences of the elderly in urban settings suggests that health literacy affects preferences for medical treatment in the context of advanced dementia.[29,30] When tools such as video decision aids are used, which are known to surmount communication and health literacy barriers, patients are more informed about their end-of-life preferences. It is unclear, however, if rural patients respond to these videos in a similar manner. To the best of our knowledge, no prior studies on end-of-life decision-making have assessed the effects of health literacy in rural populations or the use of a video decision aid to better inform rural patients.

To address this question, we conducted a randomized controlled trial of a video decision aid of the goals of care in advanced dementia among a diverse group of elderly patients in a rural community. We hypothesized that rural subjects who viewed the video would be more likely to prefer comfort-oriented measures in advanced dementia, and that subjects' preferences for level of medical care would be associated with health literacy.

Methods

Participants

Subjects were recruited from a convenience sample of patients cared for at a primary care clinic located in rural Greensburg, Louisiana, a federally designated Health Professional Shortage Area. The clinic is reflective of the village of Greensburg, which is located in St. Helena Parish. According to the Census, half of residents are African American and the other half are white, more than 99% speak exclusively English in the home and 40% of residents have less than a high school education.[31] Poverty is a significant problem for St. Helena Parish. In 2003, the per capita personal income was $19,985 compared to the state average of $26,313 and the national average of $31,472. Geographic and technological barriers are significant problems for rural populations and in particular for this community. Residents have no access to public transportation and high-speed Internet service is sparse. According to the 2000 US Census, 14% of parish residents are without telephone service.[32]

Recruitment occurred between December 1, 2008 and January 30, 2010. All scheduled English-speaking patients 65 years or older were given a flier by the clinic staff outlining the study after registering for their clinic visit, which was scheduled as part of their usual care. At the end of the visit, patients were asked by clinic staff if they were interested in participating in the study. If they indicated interest, the patient was initially interviewed by the research team for eligibility. Eligibility criteria included: (1) ability to communicate in English; (2) ability to provide informed consent; and (3) not having moderate or severe cognitive impairment (i.e., score <7) based on the Short Portable Mental Status Questionnaire (SPMSQ).[33] The protocol was approved by the Institutional Review Board of the institution affiliated with the rural clinic and all subjects provided informed consent.

Design

After obtaining informed consent, all patients meeting the eligibility criteria were randomized into a control or intervention group. The control group listened to a verbal narrative describing advanced dementia and the goals of care in advanced dementia while the intervention group listened to the same verbal narrative followed by viewing a 6-minute video visually depicting a patient with advanced dementia and the goals of care. We used simple randomization based on a computer generated scheme. All data were collected in a quiet room in the clinic area by a trained member of the research team (L.A.F.), who followed a structured script. As we have described elsewhere,[29] the interviewer read aloud the verbal narrative describing advanced dementia and the goals of care. This description was based on the Functional Assessment Staging (FAST) stage 7a,[34] which is generally considered the threshold for advanced dementia. The narrative states that advanced dementia is an incurable illness of the brain caused by many years of Alzheimer's disease or a series of strokes; its salient features are the inability to communicate understandably with others; inability to walk without assistance; and inability to feed oneself.

We then outlined three levels of medical treatments and the goals associated with each level. The first level (life-prolonging care) aims to prolong life at any cost. It includes all medically indicated treatments. The second level (limited care) aims to maintain physical functioning. It includes treatments such as hospitalization, intravenous fluids, antibiotics, but excludes attempted cardiopulmonary resuscitation and treatments in the intensive care unit. The third level (comfort care) aims to maximize comfort and to relieve pain. It includes oxygen and analgesics but excludes intravenous therapies and hospitalization unless necessary to provide comfort. Subjects were asked which level they preferred in the event they developed advanced dementia. See Appendix A for the text of the verbal description. Participants randomized to the Intervention Group viewed the video decision aid on a portable computer after listening to the same verbal narrative. The 6-minute video depicts the principal features of advanced dementia and the goals of care as described in the narrative. The video presents an 80-year-old female patient with advanced dementia together with her two daughters in the nursing home setting. The patient fails to respond to their attempts at conversation (inability to communicate). The patient is next shown being pushed in a wheelchair (inability to ambulate). Lastly, the patient is hand-fed pureed food (inability to feed oneself). Video images then followed of the goals of care in advanced dementia. Life-prolonging care images included: an intensive care unit with a ventilated patient being tended to

by respiratory therapists; a simulated code with clinicians illustrating cardiopulmonary resuscitation (CPR) and intubation; and various intravenous medications including vasopressors administered through a venous catheter. Visual images to depict limited medical care included: a patient getting antibiotics via a peripheral intravenous catheter; scenes from a typical medical ward service; and a patient wearing a nasal cannula. The video depiction of comfort care included: a patient on home hospice care receiving pain medications; a patient with a nasal cannula comfortable on oxygen at home; and, a medical attendant assisting a patient with self-care. The goals-of-care segment of the video is narrated by an African American physician. All the patients depicted in the video are white. The development of the video followed a systematic approach and has been described in detail elsewhere.[29]

Data collection and other variables

All subjects were interviewed by one member of the research team (L.A.F.) who was not blinded to randomization group. Subjects were interviewed using structured questionnaires. The baseline structured interview (15 minutes) included sociodemographic data such as age, race, gender, educational status, and marital status, as well as health status. Race was self-reported. Health status was self-rated on a Likert scale as follows: excellent, very good, good, fair, or poor. Participants were also asked if they had a diagnosis of dementia and whether they had known a person with advanced dementia.

Immediately after receiving the verbal narrative alone or narrative plus video, a second structured in-person interview (15 minutes) was conducted that included the following components: preferences for goals of care (life-prolonging care, limited care, or comfort care); health literacy; and for the intervention group, comfort using the video decision aid.

Health literacy was assessed using the Rapid Estimate of Adult Literacy in Medicine tool (REALM).[35] This is a 2- to 3-minute English test of medically relevant vocabulary. The REALM is a validated test of word pronunciation and has been shown to correlate well with tests that evaluate a range of literacy skills.[35] As others have done, we defined three categories for health literacy based on the REALM scores: sixth grade and below (REALM score of 0–45); seventh to eighth grade (REALM score of 45–60); and ninth grade and above (REALM score of 61–66).[36,37]

For those participants randomized to the video intervention group, a four-point Likert scale was used to assess perceived value of the video by asking subjects whether they had a better understanding of the disease after viewing the video, and if they would recommend the video to others.

Statistical analysis

Data were analyzed based on the decision-making modality to which each subject was randomized. The primary outcome measure was preference for care in advanced dementia ("Imagine you have advanced dementia and became very ill and in need of medical treatment. What category of care would you want to have provided: life-prolonging care, limited care, or comfort care?") categorized as three options in each of our randomization groups.

All subject characteristics and outcomes were described using proportions for categorical variables and means

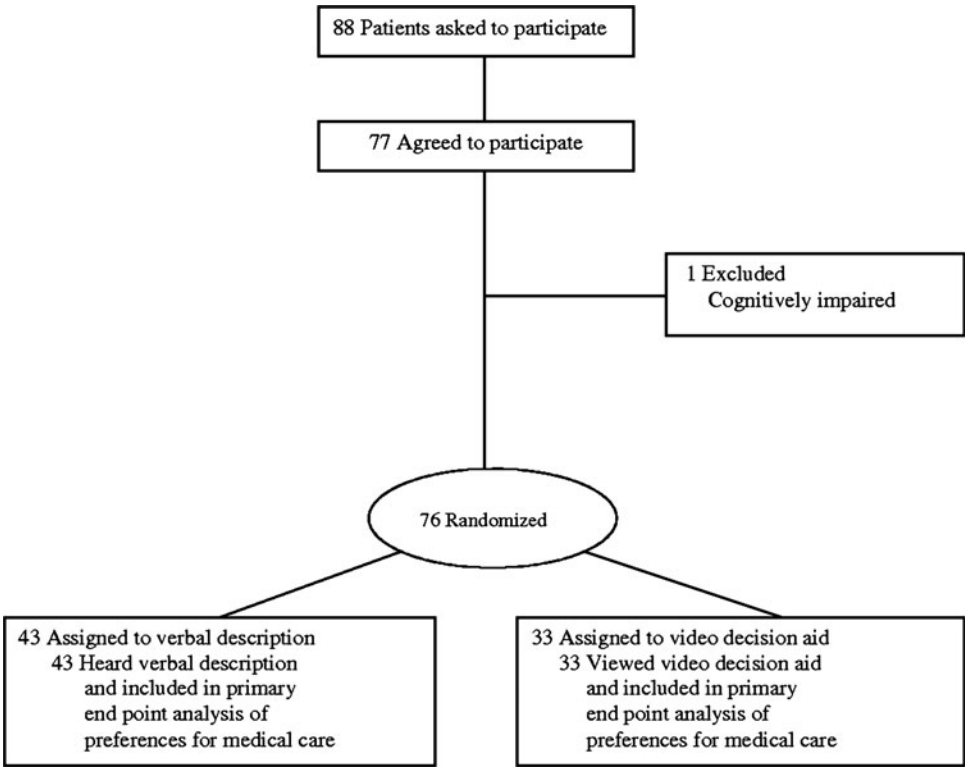


FIG. 1. Flow diagram.

(standard deviation, SD) for continuous variables. Outcomes with 95% confidence intervals (CI) were reported. Preferences for care (life-prolonging, limited, or comfort) were compared between the two groups using exact χ^2 tests.

The measure for the primary outcome analysis was the unadjusted difference in proportions of subjects preferring comfort care between the two study groups. Secondary analyses were conducted to identify factors associated with a preference for comfort care among all subjects. Bivariate analyses were conducted to determine the association between individual subject characteristics (age, gender, race, education, marital status, health status, personal history of dementia, previous relationship with a person with advanced dementia, health literacy, and randomization group) and a preference for comfort care utilizing the exact χ^2 tests. Unadjusted odds ratios (ORs) were calculated to summarize the effects of subject demographic characteristics on their preferences.

All reported p values were two-sided, with $p < 0.05$ considered statistically significant. Data were analyzed and the randomization table was prepared using SAS software, version 9.2 (SAS Institute In, Cary, NC).

Results

Study participants

A total of 88 eligible patients were approached for the study, of whom 77 (88%) agreed to be interviewed (Fig. 1). Patients who declined did not differ significantly from the recruited subjects in terms of age, gender, or race. The most common reason given for not participating was lack of interest. Of the 77 subjects recruited for the study, 1 was disqualified because her SPMSQ score was less than 7, resulting in a total of 76 study subjects. A total of 43 subjects were randomized to the control group, and 33 subjects were randomized to the video intervention group (Fig. 1). Baseline characteristics of the subjects are shown in Table 1. There were some baseline differences in the two groups including gender and marital status, but none that were statistically significant.

Outcomes

Among the 43 subjects receiving only the verbal narrative, 31 (72%) preferred comfort care, 5 (12%) chose limited care, and 7 (16%) desired life-prolonging care. Among the 33 subjects receiving the video decision aid after the verbal narrative, 30 (91%) chose comfort care, 3 (9%) chose limited care, and none desired life-prolonging care ($\chi^2 = 6.3$, $df = 2$; $p = 0.047$). Thus a significantly greater proportion of subjects in the intervention arm opted for comfort care (Fig. 2).

The unadjusted differences in proportions of subjects and odds ratios preferring comfort care as opposed to life-prolonging or limited care for each of the characteristics are listed in Table 2. Bivariate analyses revealed that factors associated with a greater likelihood to prefer comfort care among all subjects were: white race (OR 4.0; 95% CI, 1.1 to 13.9; $p = 0.041$), female (OR 3.6; 95% CI, 1.1–11.6; $p = 0.037$); randomization to the video arm (OR 3.9; 95% CI, 1.0–15.1, $p = 0.047$) and greater health literacy ($p = 0.003$). Subjects who had a health literacy level at or greater than ninth grade were more likely to have preferences for comfort care (OR 12.1; 95% CI, 2.4–62.6) compared to those with less than sixth grade level of health literacy.

The video decision support tool was highly acceptable to subjects in the intervention group: $n = 31$ of 33 (94%, 95% CI, 80–99) subjects found the video "very helpful" or "somewhat helpful," and 32 (97%, 95% CI, 84–100) said they would "definitely" or "probably" recommend the video to others. There were no adverse events in either group.

Comment

This study presents an assessment of end-of-life preferences for a group of elderly subjects from a rural community. When faced with the possibility of advanced dementia, subjects who viewed a video decision aid of advanced dementia and the goals of care were more likely to prefer comfort-oriented measures compared to patients who listened to a verbal narrative. Additionally, in unadjusted analysis, subjects were more likely to prefer a comfort-oriented approach if they had greater health literacy, were white, or were female. Finally, subjects who viewed the video found it to be very helpful, and stated that they would recommend it to other elderly patients.

To the best of our knowledge, this study represents the first randomized controlled trial looking at the end-of-life preferences of elderly patients in a rural community. Furthermore, it

Table 1. Characteristics of Community-Dwelling Elderly Persons Randomized to the Verbal Description and Video Decision Aid Groups

Characteristics	*Verbal (n = 43)*	*Video (n = 33)*
Age, mean (SD), y	75 (6)	73 (6)
Women, no. (%)	30 (70)	19 (58)
Race, no. (%)		
Black or African American	20 (47)	16 (48)
White	23 (53)	17 (52)
Health literacy,[a] no. (%)		
≤6th grade	12 (28)	12 (36)
7–8th grades	10 (23)	6 (18)
≥9th grade	21 (49)	15 (45)
Education, no. (%)		
Elementary	6 (14)	5 (15)
Some high school	12 (28)	10 (30)
High school graduate	15 (35)	6 (18)
Some college	5 (12)	5 (15)
College graduate	2 (5)	3 (9)
Postgraduate or professional	3 (7)	4 (12)
Marital status, no. (%)		
Married	14 (33)	16 (48)
Widowed	24 (56)	10 (30)
Divorced	4 (9)	6 (18)
Never married	1 (2)	1 (3)
Self-reported health status, no. (%)		
Excellent	2 (5)	3 (9)
Very good	9 (21)	11 (33)
Good	20 (47)	5 (15)
Fair	8 (19)	12 (36)
Poor	4 (9)	2 (6)
Diagnosis of dementia,[b] no. (%)	1 (2)	0
Previous relationship with a person with advanced dementia, no. (%)	8 (19)	5 (15)

[a]Health literacy assessed with the Rapid Estimate of Adult Literacy in Medicine (REALM).
[b]Subjects were asked if they had a diagnosis of dementia.
SD, standard deviation.

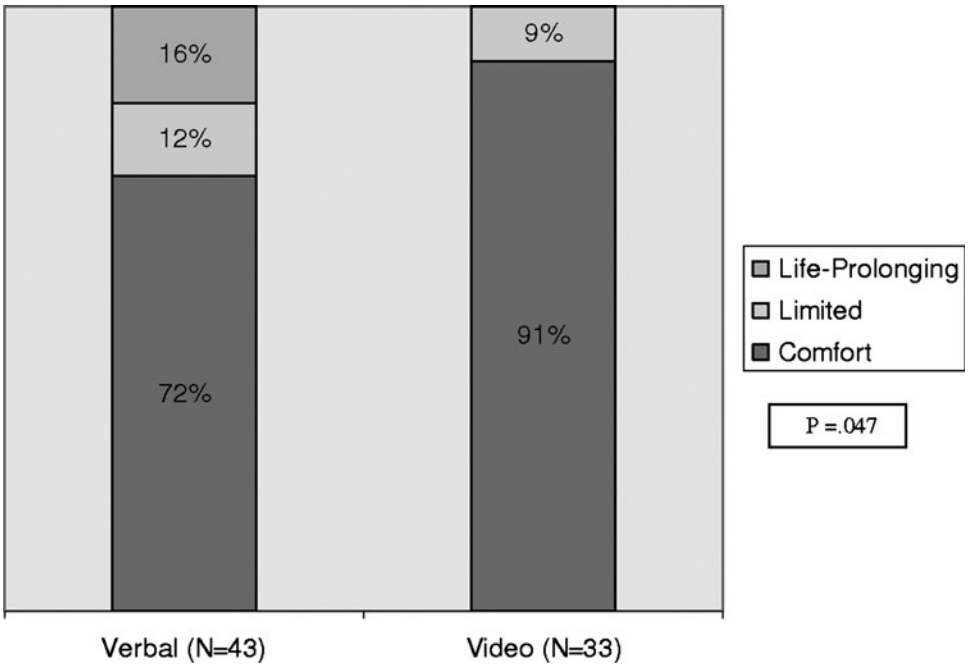


FIG. 2. Subjects' preferences for their goals of care in advanced dementia.

is the first study exploring the usefulness of a video to facilitate end-of-life discussions for elderly patients living in a rural area. Our findings are consistent with our previous investigations looking at the role of health literacy for elderly patients' preferences and the utility of a video decision aid depicting future health states to assist older subjects in expressing their preferences for end-of-life care if they were to develop advanced dementia.[29,30] The present work extends and builds upon these studies by demonstrating the role of health literacy in end-of-life preferences and the efficacy of video decision aids in end-of-life decision-making for patients in a rural setting.

Interestingly, our present study differed little from our previous randomized controlled trial of elderly subjects' preferences in an urban population.[29] While 68% of rural patients wanted comfort care in the verbal group of the present study, 64% of urban patients wanted comfort care in the verbal groups of a prior study. There was also little difference in preferences for comfort care between the video groups of the present study (91%) and our previous study (86%), although the present study included a longer video decision aid that depicted the goals of care in addition to a portrayal of advanced dementia.

In our study, white race also correlated with preferences for comfort care. This is consistent with prior studies looking at the role of race.[17–25] But this finding is limited by the fact that our study sample was not large enough to conduct multivariable analyses. Health literacy may confound the role of race. Our prior work also found an association for preferences for comfort care with white race, however, health literacy was a stronger predictor and explained the racial variance observed.[29,30] An additional association with preferring comfort care in this study was being female, which was not the case in our previous studies.[29,30]

Patients who live in rural settings are often medically underserved. Therefore, it is important to differentiate between the patients' true preferences and the type of fatalism that may be inculcated by the scarcity of medical resources. The fact that participants in our study had a similar pattern of end-of-life preferences as participants in our prior work conducted in urban settings does not exclude the possibility that improved accessibility to health services might not alter these patterns. Indeed, an appropriate critique of the current project is that we enrolled subjects at their primary care clinic, which means that our sample consists of people who already had access to health services. We did not evaluate the preferences of people outside the health setting. While this is a limitation of this study, these methods directly parallel those used in our prior projects and mimic the way we envision the video being used, i.e., in the context of a conversation with a health provider.

Additional limitations include that the interviewer collecting the data was not blinded to randomization assignment, which could have introduced bias. To minimize this potential bias, we used structured interviews and outcome measures, and a verbal script that was followed verbatim. Furthermore, prior randomized studies of interventions aimed at improving end-of-life decision-making have seldom been blinded because limiting the number of interviewers eases the burden of addressing difficult and often painful subject matter.[38–40] Second, this is a small study with a sample size of 76 subjects from one rural area in Louisiana. Thus, our findings may not be generalizable to other rural areas. We did not have the statistical power to perform additional logistic regression analyses disentangling the roles of health literacy, race and gender. Future larger studies may be of great use. Third, our video depicted white patients with an African American clinician as narrator. We did not test the robustness of our findings using other video clips that varied the features of the patient, such as race/ethnicity, gender, or clinical setting. It is possible that the race of patients or clinicians may be internalized differently for patients from different races.[41] Additional studies exploring other video clips that varied

Table 2. Univariate Analysis of the Likelihood of Choosing Comfort Care as the Primary Goal of Care

Characteristics	*Frequency in subjects choosing comfort care*	*Difference in % of subjects choosing comfort care (95% CI)*	*Univariate p Value*	*Odds ratio (95% CI)*
Age			1.0	
<80	47 (81%)			1 [Reference]
≥ 80	14 (78%)	−3% (−25% to 18%)		0.8 (0.2 to 3.0)
Gender, no. (%)			0.037	
Male	18 (67%)			1 [Reference]
Female	43 (88%)	−21% (−41% to 1%)		3.6 (1.1 to 11.6)
Education, no. (%)			1.0	
Less than college graduate	51 (80%)			1 [Reference]
College graduate or higher	10 (83%)	4% (−20% to 27%)		1.3 (0.2 to 6.5)
Marital status, no. (%)			1.0	
Not married	37 (80%)			1 [Reference]
Married	24 (80%)	0% (−19% to 18%)		1.0 (0.3 to 3.1)
Health status,[a] no. (%)			0.07	
Fair or poor	24 (92%)			1 [Reference]
Good or better	37 (74%)	−18% (−34% to −2%)		0.2 (0 to 1.1)
Diagnosis of dementia, no. (%)			1.0	
No	60 (80%)			
Yes	1 (100%)	20% (11% to 29%)		NA
Previous relationship with a person with advanced dementia, no. (%)			0.73	
No	50 (79%)			1 [Reference]
Yes	11 (84%)	5% (−17% to 27%)		1.4 (0.3 to 7.3)
Randomization, no. (%)			0.047	
Verbal	31 (72%)			1 [Reference]
Video	30 (91%)	19% (2% to 35%)		3.9 (1.0 to 15.1)
Health literacy,[b] no. (%)			0.003	
$\leq$ 6th grade	14 (58%)			1 [Reference]
7–8th grades	13 (82%)	23% (−5% to 50%)		3.1 (0.7 to 13.8)
$\geq$ 9th grade	34 (94%)	36% (15% to 57%)		12.1 (2.4–62.6)
Race, no. (%)			0.041	
Black or African American	25 (69%)			1 [Reference]
White	36 (90%)	21% (3% to 38%)		4.0 (1.1 to 13.9)

[a]Health status was one of the following: excellent, very good, good, fair, or poor.
[b]Health literacy assessed with the Rapid Estimate of Adult Literacy in Medicine (REALM).
CI, confidence interval; NA, not applicable.

features of the patients and clinician would be helpful. Finally, an emotional response to the video could have influenced subjects' preferences. To ensure that the video was not biased towards any particular perspective,[42] the video content underwent extensive scrutiny by expert physicians. Additionally, subjects' comfort level with the video is also reassuring against the possibility of a visceral reaction to the video.

Our findings underscore important similarities between the end-of-life preferences in the setting of dementia among the rural subjects in our current study and the urban subjects in our prior work. In both rural and urban settings we have observed a very high level of preference for comfort care. Furthermore, a brief educational video decision aid was associated with even a higher likelihood of choosing comfort care. This provides evidence that the differences among those who might choose comfort care, limited care, or life-prolonging care in advanced care planning settings may be due to differences in knowledge as opposed to true cultural differences. Thus, patient education and decision support is particularly warranted. These findings should be verified in further trials in a variety of settings and focusing on additional disease models.

Acknowledgments

We thank the patients and families who agreed to be filmed for this study. We also thank Jeffrey Rothschild, Chuck Pozner, and the STRATUS Center for Medical Simulation.

Dr. Volandes had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Volandes, Ferguson, Davis, Hull, Green, Deep, Paasche-Orlow. *Acquisition of data:* Ferguson. *Analysis and interpretation of data:* Volandes, Chang, Paasche-Orlow. *Drafting of the manuscript:* Volandes, Ferguson, Davis, Hull, Green, Deep, Paasche-Orlow. *Critical revision of the manuscript for important intellectual content:* Volandes, Ferguson, Davis, Hull, Green, Chang, Deep, Paasche-Orlow. *Statistical analysis:* Chang. *Obtained funding:* Volandes. *Administrative, technical, or material support:* Volandes. *Study supervision:* Volandes, Paasche-Orlow.

Dr. Volandes was supported by a George Bennett Fellowship from the Foundation for Informed Medical Decision Making and a New Investigator Research Grant from the Alzheimer's Association.

None of the Foundations participated in the collection, analysis, or interpretation of the data or in preparation, review, or approval of the manuscript.

Author Disclosure Statement

No competing financial interests exist.

References

1. U.S. Department of of Agriculture ERS. Rural population and Migration. www.ers.usda.gov/Briefing/Population/Nonmetro.htm (Last accessed May 1, 2010, 2010).
2. Robinson CA, Pesut B, Bottorff JL, Mowry A, Broughton S, Fyles G: Rural palliative care: A comprehensive review. J Palliat Med (in press).
3. Klugman CM, Dalinis PM: Ethical Issues in Rural Health Care. Baltimore: Johns Hopkins University Press, 2008.
4. Long KA, Weinert C: Rural nursing: Developing the theory base. Sch Inq Nurs Pract Summer 1989;3:113–127.
5. Waldrop D, Kirkendall AM: Rural-urban differences in end-of-life care: Implications for practice. Soc Work Health Care 2010;49:263–289.
6. Wilson DM, Justice C, Sheps S, Thomas R, Reid P, Leibovici K: Planning and providing end-of-life care in rural areas. J Rural Health 2006;22:174–181.
7. *The Quality of Medical Care in the United States: A Report on the Medicare Program: The Dartmouth Atlas of Health Care, 1999.* Chicago, IL: AHA Press, 1999.
8. Gessert CE, Haller IV: Medicare hospital charges in the last year of life: Distribution by quarter for rural and urban nursing home decedents with cognitive impairment. J Rural Health 2008;24:154–160.
9. Gessert CE, Haller IV, Kane RL, Degenholtz H: Rural-urban differences in medical care for nursing home residents with severe dementia at the end of life. J Am Geriatr Soc 2006; 54:1199–1205.
10. Levy CR, Fish R, Kramer AM: Site of death in the hospital versus nursing home of Medicare skilled nursing facility residents admitted under Medicare's Part A Benefit. J Am Geriatr Soc Aug 2004;52:1247–1254.
11. Mitchell SL, Teno JM, Roy J, Kabumoto G, Mor V: Clinical and organizational factors associated with feeding tube use among nursing home residents with advanced cognitive impairment. JAMA 2003;290:73–80.
12. Gessert CE, Mosier MC, Brown EF, Frey B: Tube feeding in nursing home residents with severe and irreversible cognitive impairment. J Am Geriatr Soc 2000;48:1593–1600.
13. Gessert CE, Calkins DR: Rural-urban differences in end-of-life care: The use of feeding tubes. J Rural Health Winter 2001;17:16–24.
14. Campbell CL, Merwin E, Yan G: Factors that influence the presence of a hospice in a rural community. J Nurs Scholarsh 2009;41:420–428.
15. Menec VH, Nowicki S, Kalischuk A: Transfers to acute care hospitals at the end of life: Do rural/remote regions differ from urban regions? Rural Remote Health 2010;10:1281.
16. Madigan EA, Wiencek CA, Vander Schrier AL: Patterns of community-based end-of-life care in rural areas of the United States. Policy Polit Nurs Pract 2009;10:71–81.
17. Garrett JM, Harris RP, Norburn JK, Patrick DL, Danis M: Life-sustaining treatments during terminal illness: Who wants what? J Gen Intern Med 1993;8:361–368.
18. Caralis PV, Davis B, Wright K, Marcial E: The influence of ethnicity and race on attitudes toward advance directives, life-prolonging treatments, and euthanasia. J Clin Ethics 1993;4:155–165.
19. O'Brien LA, Grisso JA, Maislin G, LaPann K, Krotki KP, Greco PJ, Siegert EA, Evans LK: Nursing home residents' preferences for life-sustaining treatments. JAMA 1995; 274:1775–1779.
20. Blackhall LJ, Frank G, Murphy ST, Michel V, Palmer JM, Azen SP: Ethnicity and attitudes towards life sustaining technology. Soc Sci Med 1999;48:1779–1789.
21. Crawley L, Payne R, Bolden J, Payne T, Washington P, Williams S: Palliative and end-of-life care in the African American community. JAMA 2000;284:2518–2521.
22. Hopp FP, Duffy SA: Racial variations in end-of-life care. J Am Geriatr Soc 2000;48:658–663.
23. Johnson KS, Elbert-Avila KI, Tulsky JA: The influence of spiritual beliefs and practices on the treatment preferences of African Americans: A review of the literature. J Am Geriatr Soc 2005;53:711–719.
24. Sehgal A, Galbraith A, Chesney M, Schoenfeld P, Charles G, Lo B: How strictly do dialysis patients want their advance directives followed? JAMA 1992;267:59–63.
25. Klessig J: The effect of values and culture on life-support decisions. West J Med 1992;157:316–322.
26. Kleinman A: *The Illness Narratives: Suffering, Healing, And The Human Condition.* New York: Basic Books, 1988.
27. Payer L: *Medicine & Culture: Varieties of Treatment in the United States, England, West Germany, and France, 1st ed.* New York: H. Holt, 1988.
28. Hern HE, Jr., Koenig BA, Moore LJ, Marshall PA: The difference that culture can make in end-of-life decisionmaking. Camb Q Healthc Ethics 1998;7:27–40.
29. Volandes AE, Paasche-Orlow MK, Barry MJ, Gillick MR, Minaker KL, Chang Y, Cook EF, Abbo ED, El-Jawahri A, Mitchell SL: Video decision support tool for advance care planning in dementia: Randomised controlled trial. BMJ 2009;338:b2159.
30. Volandes AE, Paasche-Orlow M, Gillick MR, Cook EF, Shaykevich S, Abbo ED, Lehmann L: Health literacy not race predicts end-of-life care preferences. J Palliat Med 2008;11: 754–762.
31. National Center for Education Statistics N: www.casas.org/lit/litcode/Detail.CFM?census__AREAID=1505 (Last accessed July 13, 2010).
32. St. Helena Parish Health Profile D. http://dhh.louisiana.gov/OPH/PHP%202005/PDF/St.%20Helena/Economics%20St%20Helena.pdf (Last accessed July 16, 2010).
33. Pfeiffer E: A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc 1975;23:433–441.
34. Sclan SG, Reisberg B: Functional assessment staging (FAST) in Alzheimer's disease: Reliability, validity, and ordinality. Int Psychogeriatr 1992;4(Suppl 1):55–69.
35. Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, Crouch MA: Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med 1993;25:391–395.
36. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR: The prevalence of limited health literacy. J Gen Intern Med 2005;20:175–184.

37. Davis TC, Wolf MS, Bass PF, 3rd, Thompson JA, Tilson HH, Neuberger M, Parker RM: Literacy and misunderstanding prescription drug labels. Ann Intern Med 2006;145:887–894.
38. SUPPORT: A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). The SUPPORT Principal Investigators. JAMA 1995;274:1591–1598.
39. Hofmann JC, Wenger NS, Davis RB, Teno J, Connors AF Jr, Desbiens N, Lynn J, Phillips RS: Patient preferences for communication with physicians about end-of-life decisions. SUPPORT Investigators. Study to Understand Prognoses and Preference for Outcomes and Risks of Treatment. Ann Intern Med 1997;127:1–12.
40. Covinsky KE, Fuller JD, Yaffe K, Johnston CB, Hamel MB, Lynn J, Teno JM, Phillips RS: Communication and decision-making in seriously ill patients: findings of the SUPPORT project. The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. J Am Geriatr Soc 2000;48(5 Suppl):S187–193.
41. Jibaja-Weiss ML, Volk RJ: Utilizing computerized entertainment education in the development of decision aids for lower literate and naive computer users. J Health Commun 2007;12:681–697.
42. Winterbottom A, Bekker HL, Conner M, Mooney A: Does narrative information bias individual's decision making? A systematic review. Soc Sci Med 2008;67:2079–2088.

Address correspondence to:
Angelo E. Volandes, M.D.
Massachusetts General Hospital
General Medicine Unit
50 Staniford Street, 9th Floor
Boston, MA 02114

E-mail: avolandes@partners.org

(*Appendix follows* →)

Appendix A. Narrative Describing Advanced Dementia

I am going to describe to you an illness called advanced dementia, like advanced Alzheimer's dementia, that you may or may not be familiar with. Advanced dementia is an incurable disease of the brain in which one is not able to communicate with others. People in advanced dementia are not able to move around or walk, get out of bed independently, eat by oneself, or communicate understandably with others. People with advanced dementia often have difficulty chewing or swallowing, and require assistance with feeding oneself. Advanced dementia is an incurable disease and most commonly occurs after many years of Alzheimer's disease or as the result of strokes. People are not able to answer any questions or tell you about themselves.

Narrative Describing the Goals of Care

I am going to ask you a question about your preferences for medical care if you had a disease called advanced dementia. I will ask you what you prefer. You have three choices for medical care if you had this condition. I will first review these three choices with you. The three choices for medical care that I want you to think about for advanced dementia are life-prolonging care, limited care, and comfort care.

Life-Prolonging Care

The goal of this category of care is to prolong life. There are no limits to care. This choice includes everything a modern hospital has to offer to maintain your life. Such procedures include: cardiopulmonary resuscitation or CPR in which a doctor pushes on your chest when the heart stops and will often use electricity to shock the heart; being placed on a breathing machine, also known as life support, in which a tube is placed down your throat into the lungs; and other medical procedures performed in the intensive care unit or ICU. The goal is to prolong life.

Limited Care

The goal of this category is to maintain physical and mental functions. Care will depend on your physical and mental functioning. Such care includes intravenous (IV) therapies like antibiotics and hospitalization. But does not include cardiopulmonary resuscitation/CPR and intensive care unit/ICU care. The goal is to maintain physical and mental functioning.

Comfort Care

The goal of this category is to maximize comfort. Only measures that comfort or relieve pain are performed. The aim is to relieve pain and to be kept as pain-free as possible. Comfort care does not include cardiopulmonary resuscitation/CPR, respirators, intensive care unit/ICU care, and generally would not include IV therapy or hospitalization. The goal is maximizing comfort and relieving pain.

Imagine you have advanced dementia and became very ill and in need of medical treatment. What category of care would you want to have provided: Life-Prolonging Care, Limited Care, or Comfort Care?

ELSEVIER

Neurobiology of Aging 32 (2011) 2109.e15–2109.e28

NEUROBIOLOGY OF AGING

www.elsevier.com/locate/neuaging

A genome-wide association study of aging

Stefan Walter[a,b,1], Gil Atzmon[c,d,e,1], Ellen W. Demerath[f,1], Melissa E. Garcia[g,1], Robert C. Kaplan[h,1], Meena Kumari[i,1], Kathryn L. Lunetta[j,1], Yuri Milaneschi[k,1], Toshiko Tanaka[k,1], Gregory J. Tranah[l,1], Uwe Völker[m,1], Lei Yu[n,1], Alice Arnold[o], Emelia J. Benjamin[p,q], Reiner Biffar[r], Aron S. Buchman[n], Eric Boerwinkle[s], David Couper[t], Philip L. De Jager[u], Denis A. Evans[v], Tamara B. Harris[g], Wolfgang Hoffmann[w,x], Albert Hofman[b], David Karasik[y], Douglas P. Kiel[y], Thomas Kocher[r], Maris Kuningas[b], Lenore J. Launer[g], Kurt K. Lohman[z], Pamela L. Lutsey[f], Johan Mackenbach[a], Kristin Marciante[aa], Bruce M. Psaty[aa,bb], Eric M. Reiman[cc], Jerome I. Rotter[dd], Sudha Seshadri[p,q], Michelle D. Shardell[ee], Albert V. Smith[ff], Cornelia van Duijn[b], Jeremy Walston[gg], M. Carola Zillikens[hh], Stefania Bandinelli[ii,2], Sebastian E. Baumeister[w,2], David A. Bennett[n,2], Luigi Ferrucci[jj,2], Vilmundur Gudnason[ff,2], Mika Kivimaki[i,2], Yongmei Liu[z,2], Joanne M. Murabito[p,q,2], Anne B. Newman[kk,2], Henning Tiemeier[b,ll,2,*], Nora Franceschini[mm,2]

[a] *Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands*
[b] *Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands*
[c] *Institute for Aging Research and the Diabetes Research Center, Albert Einstein College of Medicine, Bronx, NY, USA*
[d] *Department of Medicine, Albert Einstein College of Medicine, Bronx, NY, USA*
[e] *Department of Genetic, Albert Einstein College of Medicine, Bronx, NY, USA*
[f] *Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA*
[g] *Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA*
[h] *Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA*
[i] *Department of Epidemiology and Public Health, University College London, London, UK*
[j] *Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA*
[k] *Clinical Research Branch, National Institute on Aging, Baltimore, MD, USA*
[l] *California Pacific Medical Center, San Francisco, CA, USA*
[m] *Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany*
[n] *Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA*
[o] *Department of Biostatistics, University of Washington, Seattle, WA, USA*
[p] *Sections of General Internal Medicine, Preventive Medicine, Cardiology, and Neurology, Department of Medicine, Boston University School of Medicine, Boston, MA, USA*
[q] *The National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, MA, USA*
[r] *Dental School, Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany*
[s] *Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA*
[t] *Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA*
[u] *Program in Translational NeuroPsychiatric Genomics, Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*
[v] *Rush Institute for Healthy Aging, Rush University Medical Center, Chicago, IL, USA*
[w] *Institute of Community Medicine, Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany*
[x] *Center for Integrated Dementia Care Research (CIDC), a scientific cooperation between the Universities and University Hospitals of Rostock and Greifswald, and the German Center for Neurodegenerative Disease (DZNE), Bonn, Germany*
[y] *Institute for Aging Research, Hebrew SeniorLife and Harvard Medical School, Boston, MA, USA*
[z] *Center for Human Genomics, Department of Epidemiology and Prevention, Wake Forest University School of Medicine, Winston-Salem, NC, USA*
[aa] *Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, WA, USA*
[bb] *Group Health Research Unit, Group Health Cooperative, Seattle, WA, USA*
[cc] *Neurogenomics Division, The Translational Genomics Research Institute, Banner Alzheimer's Institute, Phoenix, AZ, USA*
[dd] *Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA*

doi:10.1016/j.neurobiolaging.2011.05.026

[ee] Epidemiology and Public Health, University of Maryland, MD, USA
[ff] Icelandic Heart Association, Kópavogur, Iceland
[gg] Johns Hopkins University School of Medicine, Division of Geriatric Medicine and Gerontology, Baltimore, MD, USA
[hh] Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands
[ii] Geriatric Unit, Azienda Sanitaria Firenze (ASF), Florence, Italy
[jj] Clinical Research Branch, National Institute on Aging, Baltimore, MD, USA
[kk] Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA
[ll] Department of Child and Adolescent Psychiatry, Erasmus Medical Center, Rotterdam, The Netherlands
[mm] Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

Received 7 January 2011; received in revised form 14 April 2011; accepted 30 May 2011

Abstract

Human longevity and healthy aging show moderate heritability (20%–50%). We conducted a meta-analysis of genome-wide association studies from 9 studies from the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium for 2 outcomes: (1) all-cause mortality, and (2) survival free of major disease or death. No single nucleotide polymorphism (SNP) was a genome-wide significant predictor of either outcome ($p < 5 \times 10^{-8}$). We found 14 independent SNPs that predicted risk of death, and 8 SNPs that predicted event-free survival ($p < 10^{-5}$). These SNPs are in or near genes that are highly expressed in the brain (*HECW2*, *HIP1*, *BIN2*, *GRIA1*), genes involved in neural development and function (*KCNQ4*, *LMO4*, *GRIA1*, *NETO1*) and autophagy (*ATG4C*), and genes that are associated with risk of various diseases including cancer and Alzheimer's disease. In addition to considerable overlap between the traits, pathway and network analysis corroborated these findings. These findings indicate that variation in genes involved in neurological processes may be an important factor in regulating aging free of major disease and achieving longevity.

Keywords: Genome-wide association analysis; Mortality; Disease-free survival; Longevity; Aging; Brain aging

1. Introduction

The recent, remarkable extension of life expectancy is largely attributed to the postponement of mortality at old age (Vaupel, 1997, 2010). The years of life gained in the older population residing in developed nations are a success story of public health measures and improved health care. In addition to such external factors, longevity and healthy aging consistently show a modest heritability between 20% and 50% and aging-associated genetic research may provide further insights into the mechanisms of aging (Herskind et al., 1996; McGue et al., 1993; Reed and Dick, 2003). It has been postulated that genes involved in pathways associated with aging identified in animal models, such as insulin-like growth factor (IGF)-insulin signaling, regulation of lipoprotein metabolism, the mTOR pathway, and the oxidative stress response may also influence survival to old or even exceptionally old age in humans (Christensen et al., 2006; Kenyon, 2010; Vellai et al., 2003). However, in humans, common variants within genes involved in these pathways have not been consistently associated with lifespan (Christensen et al., 2006; Kenyon, 2010; Kuningas et al., 2008; Vijg and Suh, 2005).

The lack of success in the identification of genes related to aging in humans may be due to the complexity of the phenotype. One approach to investigate aging and longevity is to compare frequencies of genetic variants between nonagenarians or centenarians and the general population. This approach led to the discovery of an association between *APOE* (Deelen et al., 2011; Ewbank, 2007; Gerdes et al., 2000) and more recently *FOXO3A* (Anselmi et al., 2009; Flachsbart et al., 2009; Li et al., 2009a; Pawlikowska et al., 2009; Willcox et al., 2008) and human aging and longevity. However, a recent genome-wide association study (GWAS) of individuals reaching the age of 90 or older failed to identify genome-wide significant variants (Newman et al., 2010).

Prospective follow-up studies with a continuous outcome such as time to death are more powerful than case-control analyses. A study of time to death simultaneously addresses the effects of genetic variation related to life span, the progression toward death, and disease-specific mortality. This design has been successfully applied in animal models (Finch and Ruvkun, 2001; Kenyon, 2010) and also in human genetics research of blood pressure (Levy et al., 2009; Newton-Cheh et al., 2009; van Rijn et al., 2007), a trait with heritability similar to longevity, where examination of a continuous outcome has been more successful in identifying genetic loci than studies that have solely used hypertension

[1] These authors contributed equally to this work.
[2] These authors are joint senior authors on this work.
* Corresponding author at: Erasmus Medical Center Rotterdam, Department of Epidemiology & Biostatistics, P.O. Box 1738, 3000 DR Rotterdam, Netherlands. Tel.: +31 10 70 32183; fax: +31 10 70 44657. (H. Tiemeier).
E-mail address: h.tiemeier@erasmusmc.nl (H. Tiemeier).

as a dichotomous trait. Frailty and survival free of disease have been suggested as more promising phenotypes for studies of aging because mortality is a very heterogeneous outcome caused by multiple chronic conditions (Vijg and Suh, 2005).

This study addresses the genetics of aging in a broad, sequential way using data from cohort studies participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. First, we aimed to identify single nucleotide polymorphism (SNPs) associated with all cause mortality (time to death) in a hypothesis-free GWAS in approximately 25,000 unselected persons of European ancestry. Second, we performed GWAS of time to event, defined by major incident events (myocardial infarction, heart failure, stroke, dementia, hip fracture, or cancer) or death, as an alternative phenotype for healthy aging. Last, we analyzed the SNPs along with their respective most likely associated genes identified in the GWAS meta-analyses to identify pathways and networks associated with aging and longevity.

2. Methods

2.1. Participants

The participants are of recent European ancestry and stem from cohorts of the CHARGE Consortium (Psaty et al., 2009). All cohorts are follow-up studies periodically assessing the health and vital status of their participants. Although some of the cohorts included multiple ethnic groups, only data from self-reported Caucasians were used. In addition, population structure was assessed using principal components in each CHARGE study and outliers were removed. Any remaining within-study structure was adjusted for using appropriate methods (Price et al., 2006). All participants included in this analysis were at least 55 years of age at the time of blood draw for DNA and provided written informed consent. A brief description of each population is given in the Supplementary Information.

2.2. Phenotype

We conducted a survival analysis, adjusted for age at baseline and sex, to model continuous time to death or end of follow-up in 25,007 participants (deceased "cases" = 8444; mean follow-up time = 10.6 [SD 5.4] years) that were older than 55 years at baseline. As research demonstrated that the likelihood of incident disease is genetically determined, we defined a second phenotype: survival free of major disease or mortality (Atzmon et al., 2004; Lunetta et al., 2007; Vijg and Suh, 2005). The outcome was defined as time to the first of the following adjudicated events: myocardial infarction, heart failure, stroke, dementia, hip fracture, cancer, or death. For this analysis, participants at baseline were older than 55 years of age and free of any of the aforementioned conditions. Inclusion in the study required complete follow-up information on mortality and at least 4 out of 6 of the health conditions. Genome-wide information on polymorphisms was available for 16,995 participants free of disease at the beginning of the study. These participants were followed for 8.8 (SD 5.7) years and we registered 7314 major events.

2.3. Genotyping and imputation

As different genotyping platforms were used across studies, we imputed to 2.5 million SNPs using the HapMap 22 CEU (Build 36) genotyped samples as a reference. For details on the study-specific quality control procedures for genotyping and imputation please consult Supplementary Table S1.

2.4. Statistical analysis

We used the semiparametric Cox proportional hazard to model time to event for both phenotypes in each study. Follow-up time since baseline was used as time scale. An additive genetic model was used in this analysis. We subsequently combined the individual study results in a meta-analysis using a fixed effects model that combined the study-specific regression parameters and standard errors using inverse variance weighting. We included SNPs that had a minor allele frequency (MAF) of at least 1% and an imputation quality ratio (de Bakker et al., 2008) (equivalent to the MaCH r^2 statistic; Li et al., 2009b) of at least 0.3. The study-specific inflation factors (λ_{GC}) were computed using the set of chi-square statistics used for the meta-analysis for each study. The inflation factor is computed as the median of all chi-square statistics divided by the expected median of the statistics (approximately 0.456) for a chi-square distribution with 1 degree of freedom. SNP associations at $p < 5 \times 10^{-8}$ were considered to be genome-wide significant. SNPs with $p < 5 \times 10^{-5}$ were considered suggestive associations. The combined meta-analysis hazard ratio (HR) can be interpreted as the increase in the risk of dying or having a major event during follow-up per additional copy of the coded allele. Power analysis revealed 80% statistical power to detect SNPs with a minor allele frequency of 5% and relative risk of 1.10 using a nominal significance level of 0.05 (Supplementary Table S2).

In addition, we incorporated gene annotation information, a technique that has successfully been applied in the field of aging research (de Magalhaes et al., 2009a, 2010). Protein ANalysis THrough Evolutionary Relationships (PANTHER; Mi et al., 2007; Thomas et al., 2003) and Ingenuity Pathway Analysis (IPA) (www.ingenuity.com) were used for identification and classification of networks, pathways, biological processes, and molecular functions of the genes identified in this study. For both phenotypes we generated lists of candidate genes. These genes were the closest reference genes to the SNPs associated with the outcome at $p < 1 \times 10^{-3}$. PANTHER compares these gene lists to the reference list using the binomial test for each molecular function, biological process, or pathway term.

IPA builds networks by searching the Ingenuity Pathways Knowledge Base for interactions between the identified genes and all other gene objects stored in the knowledge base.

3. Results

We conducted a meta-analysis of GWAS on time to death adjusted for baseline age and sex in participants of European origin, 55 years of age or older from 9 longitudinal cohort studies participating in the CHARGE Consortium (Psaty et al., 2009). In total, we observed 8444 deaths (mean age at death: 81.1, SD 8.4) in 25,007 participants (55% female) after an average follow-up of 10.6 (SD 5.4) years. Descriptive characteristics of participants and Manhattan plots showing genome wide *p*-values for association are displayed in the Supplementary data (Supplementary Fig. S1, and Supplementary Tables S3 and S4). The quantile-quantile plot (Q-Q plot) of observed versus expected *p*-values showed only a small deviation from the null hypothesis, indicating no significant population stratification (Fig. 1a, $\lambda_{GC} = 1.066$). Although there were no genome-wide significant findings ($p < 5 \times 10^{-8}$), 14 independent SNPs were associated with time to death at a suggestive threshold of $p < 1 \times 10^{-5}$ (Table 1). Among these SNPs, rs4936894 (chromosome 11, near the von Willebrand factor A domain containing 5A gene [*VWA5A*]) had the strongest association with time to death ($p = 3.4 \times 10^{-7}$). We sought replication for 5 of the 14 top SNPs with the strongest association with time to death in 4 independent samples ($n = 10{,}411$, deaths = 1295) of the same ancestry. None of the SNPs were consistently associated with time to death at a nominally significant level of $p < 0.05$ across all replication samples (Supplementary Tables S5–S8). In the combined meta-analysis of the discovery and replication studies only rs1425609 in the vicinity of otolin-1 (*OTOL1*) showed a stronger association (1.61×10^{-6}).

Likewise, no genome-wide significant findings were identified in the time to event analysis following 16,995 participants free of disease at baseline and registering 7314 events over an average of 8.8 (SD 5.7) years of follow-up (Table 2). Events included incident myocardial infarction, heart failure, stroke, dementia, hip fracture, and cancer or death. The Q-Q plot (Fig. 1a, $\lambda_{GC} = 1.019$) showed no evidence of inflation of type I error. In total, there were 8 independent SNPs associated with event-free survival at $p < 10^{-5}$. The SNP with the strongest association was rs10412199 (chromosome 19, $p = 3.02 \times 10^{-6}$), which is in close proximity to ataxia, cerebellar, Cayman type (*ATCAY*). Additional descriptive information including definitions of each event and association results with $p < 10^{-4}$ are provided in Supplementary Figure S2, and Supplementary Tables S9–S12.

As both phenotypes may provide different but complimentary information about the aging process, we evaluated the overlap between their association results (Table 3). Interpretation of the overlap between the phenotypes requires caution as both phenotypes are correlated, nevertheless it helps to focus on specific loci and put them into the context of aging. From the 14 loci passing the prespecified, suggestive threshold of $p < 1 \times 10^{-5}$ in the time to death analysis, 5 had corresponding SNPs within 500 kilo base pairs distance, in linkage disequilibrium (LD; $r^2 > 0.1$) associated with $p < 1 \times 10^{-4}$ and the same overall direction of the effect in the time to event analysis. These 5 regions were in the vicinity of the following genes: *OTOL1* (3q26.1), bridging integrator 2 (*BIN2*, 12q13), ATG4 autophagy related 4 homolog C (*ATG4C*, 1p31.3), origin recognition complex,

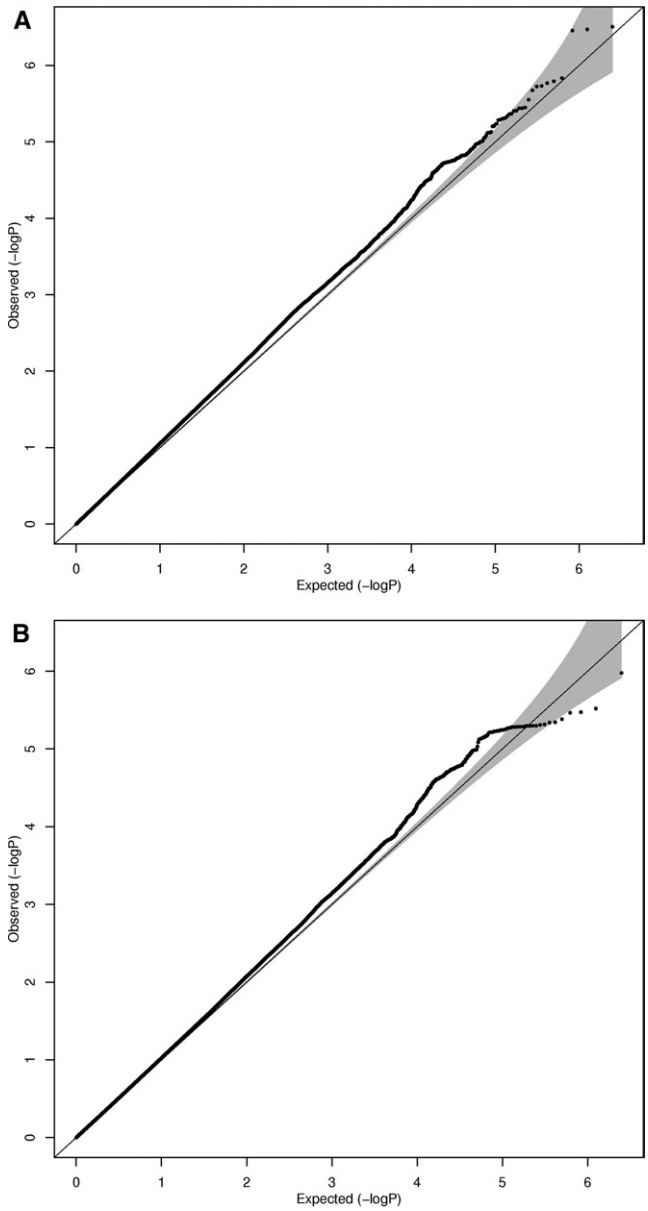


Fig. 1. (a) Quantile-quantile (Q-Q) plot after meta-analysis for time to death. (b) Quantile-quantile (Q-Q) plot after meta-analysis for time to event.

Table 1
Top 14 SNPs (p-value $< 10^{-5}$) for time to death ranked by p-value, from meta-analysis of 9 cohorts[a]

Number	SNP	Chr	Position	Closest reference gene	Distance from closest gene	Coded allele	Noncoded allele	Frequency coded allele	HR	p-value	Study effect direction [b]	Number of supporting SNPs
1	rs4936894	11	123522703	*VWA5A*	123	A	G	0.226	1.11	3.38E-07	++++-++-+	224
2	rs1425609	3	164164689	*OTOL1*	1,460,265	A	G	0.381	0.92	1.46E-06	—-------	399
3	rs766903	12	49990101	*BIN2*	14,104	A	G	0.834	0.90	1.61E-06	———————+----	7
4	rs12042640	1	63139384	*ATG4C*	36,747	T	C	0.284	1.09	1.71E-06	++++-+-+-	19
5	rs17149227	7	75073485	*HIP1*	72,141	T	G	0.959	0.79	3.56E-06	−??———————+-?	0
6	rs3128591	9	136741940	*COL5A1*	68,468	A	G	0.754	0.92	3.64E-06	—-------	20
7	rs11582903	1	87618642	*LMO4*	34,804	A	C	0.150	1.12	3.94E-06	++-+++++	38
8	rs4850695	2	196861504	*HECW2*	89,283	A	G	0.766	1.09	4.62E-06	+++++++++	95
9	rs10259086	7	103680248	*ORC5L*	44,549	T	G	0.686	1.08	5.16E-06	++++++-++	72
10	rs2769255	1	41017941	*KCNQ4*	4329	T	C	0.374	1.08	5.17E-06	++++++-++	95
11	rs17291546	6	2660681	*LOC340156*	35,472	A	G	0.957	0.82	7.65E-06	−?-------	8
12	rs12606100	18	69102967	*NETO1*	417,177	T	C	0.202	1.11	8.72E-06	+??-++++-	4
13	rs1274214	11	122979741	*GRAMD1B*	18,987	T	C	0.500	0.93	8.87E-06	—-------	42
14	rs10811679	9	2224701	*SMARCA2*	41,080	T	C	0.330	1.08	9.53E-06	+++++++++	37

n = 25,007 participants with 8444 deaths, only SNPs with MAF > 3% are presented. p-values are for the inverse variance-weighted meta-analysis. Distances to genes are given in base pairs. Position is for NCBI Build 36. HRs are for each additional coded allele. Number of supporting SNPs is the number of SNPs within 500 kilo base pairs of the top SNP that are in LD with the top SNP in the HapMap CEU release 22 ($r^2 >= 0.10$) and have association p-value < 0.05.

Key: Chr, chromosome; LD, linkage disequilibrium; MAF, minor allele frequency; SNP, single nucleotide polymorphism.

[a] For information on all SNP associations with p-value $< 10^{-4}$ see Supplementary Table S2.

[b] Study-specific information is presented in the order: RS, CHS, FHS, ARIC, AGES, HABC, BLSA, InCHIANTI, SHIP; "+" = coded allele increases risk of mortality, "-" = coded allele decreases risk of mortality, "?" = not tested.

Table 2
Top 8 SNP (p-value $< 10^{-5}$) associations from meta-analysis of 8 cohorts for time to event, ranked by p-value (n = 16,995 with 7314 events)

Number	SNP	Chr	Position	Closest reference Gene	Distance (bp) from closest gene	Coded allele	Noncoded allele	Frequency coded allele	HR	p-value	Study effect direction [a]	Number of supporting SNPs
1	rs10412199	19	3878771	ATCAY	307	A	G	0.33	0.91	3.02E-06	–?-++--	6
2	rs16852912	3	170169370	MECOM	114610	T	C	0.08	1.18	3.37E-06	++++-+-+	72
3	rs8001976	13	47285723	SUCLA2	129069	T	C	0.44	1.09	3.43E-06	+++-+++--	173
4	rs11162963	1	80507169	ELTD1	1262086	T	C	0.63	1.09	4.15E-06	+++++++++	40
5	rs4764043	12	14006749	GRIN2B	17570	T	C	0.08	1.17	6.10E-06	+++++++++	2
6	rs3112530	5	152619870	GRIA1	230628	A	G	0.08	0.85	6.79E-06	----------+---	130
7	rs10202497	2	237935633	COL6A3	38233	A	C	0.14	0.89	8.22E-06	----------	36
8	rs2367725	1	43988415	ST3GAL3	42611	T	C	0.42	1.08	9.31E-06	+++++++-	119

p-values are for the inverse variance-weighted meta-analysis. Distances to genes are given in base pairs. Position is for NCBI Build 36. HRs are for each additional coded allele. Number of supporting SNPs is the number of SNPs within 500 kilo base pair of the top SNP that are in LD with the top SNP in the HapMap CEU release 22 ($r^2 \geq 0.10$) and have association p-value < 0.05. For information on all SNP associations with p-value $< 10^{-4}$ see Supplementary Table S12.
Key: bp, base pair; Chr, chromosome; HR, hazard ratio; LD, linkage disequilibrium; SNP, single nucleotide polymorphism.
[a] Study-specific information is presented in the order: RS, CHS, FHS, ARIC, AGES, HABC, BLSA, InCHIANTI; "+" = coded allele increases risk of event; "-" = coded allele decreases risk of event; "?" = not tested.

subunit 5-like (*ORC5L*, 7q22.1), and potassium voltage-gated channel, KQT-like subfamily, member 4 (*KCNQ4*, 1p34). Similarly, in the time to event analysis 3 of the 8 top SNPs showed considerable overlap and the same direction of effect in the time to death analysis. These SNPs were close to the following genes: MDS1 and EVI1 complex locus (*MECOM*, 3q24-q28), succinate-CoA ligase, ADP-forming, beta subunit (*SUCLA2*, 13q12.2-q13.3), and ST3 beta-galactoside alpha-2,3-sialyltransferase 3 (*ST3GAL3*, 1p34.1).

Finally, we evaluated candidate genes for aging by identification and classification of networks, pathways, biological processes, and molecular functions. The candidate genes were derived from the meta-analyses of GWAS and included the reference genes closest to the SNPs associated with $p < 1 \times 10^{-3}$ (time to death: 862 genes, time to event: 704 genes). We used PANTHER (Mi et al., 2007; Thomas et al., 2003, 2006) and IPA software (www.ingenuity.com) for these analyses. PANTHER compares these gene lists to the reference list using the binomial test for each molecular function, biological process, or pathway term. IPA builds networks by searching the Ingenuity Pathways Knowledge Base for interactions between the identified genes and all other gene objects stored in the knowledge base.

For the analysis of time to death, the relevant biological processes overrepresented in the PANTHER analysis were developmental processes, neuronal activities, signal transduction, neurogenesis, ectoderm development, and cell adhesion. For the analysis of time to incident event, developmental processes and neuronal activities were overrepresented among other biological process (Table 4). The analyses also highlighted the Wnt signaling pathway. The Wnt signaling pathway is ubiquitous and known to be involved in cancer but also plays an important role in the early stages of the development of the central nervous system, in synaptic formation by axon guidance, and in modulating fibrosis during muscle repair scored high in both traits under study (Brack et al., 2007; Inestrosa and Arenas, 2010; Keeble et al., 2006; Ulloa and Martí, 2010). For extended tables see Supplementary Tables S13 and S14. In addition, Ingenuity identified 1 network with $p = 10^{-31}$ containing 26 genes involved in processes related to nervous system development and function for the analysis of time to death (Fig. 2) and 1 network with $p = 10^{-40}$ containing 28 genes involved in cellular function and development for time to event (Supplementary Fig. S3).

IPA analysis highlighted the following genes associated with the time to death trait: *NTRK2* (neurotrophic tyrosine kinase, receptor, type 2)—a member of the neurotrophic tyrosine receptor kinase family. This kinase is a membrane-bound receptor that, upon neurotrophin binding, phosphorylates itself and members of the mitogen-activated protein kinase (MAPK) pathway. Signaling through this kinase leads to cell differentiation. Second in line were *NCAM1* (neural cell adhesion molecule 1)—a cytoskeletal binding

Table 3
Overlap between the associations of time to death and time to event[a]

Top hit	SNP	Chr	Closest reference gene	Time to death		Time to event		Top SNPs from time to death (time to event) analysis associated with different *p*-values in time to event (time to death) analysis					
				p	Effect	*p*	Effect	Total	$p \leq 0.05$	$p < 0.05$	$p < 0.01$	$p < 0.001$	$p < 0.0001$
Time to death													
1	rs1425609	3	*OTOL1*	1.46E-06	−	0.005704	−	1119	693	235	132	37	22
2	rs766903	12	*BIN2*	1.61E-06	−	0.01315	−	37	27	4	5	0	1
3	rs12042640	1	*ATG4C*	1.71E-06	+	0.03701	+	93	60	19	4	0	10
4	rs11582903	1	*LMO4*	3.94E-06	+	0.7336	−	133	91	8	12	21	1
5	rs10259086	7	*ORC5L*	5.16E-06	+	0.03266	+	239	154	56	21	4	4
6	rs2769255	1	*KCNQ4*	5.17E-06	+	0.01322	+	287	151	68	56	7	5
7	rs17291546	6	*LOC340156*	7.65E-06	−	0.01624	−	29	19	9	1	0	0
8	rs12606100	18	*NETO1*	8.72E-06	+	0.02853	+	23	16	5	2	0	0
9	rs1274214	11	*GRAMD1B*	8.87E-06	−	0.0567	−	101	39	28	17	17	0
Time to event													
1	rs16852912	3	*MECOM*	0.00589	+	3.37E-06	+	169	67	49	49	2	2
2	rs8001976	13	*SUCLA2*	0.01473	+	3.43E-06	+	433	198	91	46	59	39
3	rs4764043	12	*GRIN2B*	0.0017	+	6.10E-06	+	45	42	2	1	0	0
4	rs10202497	2	*COL6A3*	0.00035	−	8.22E-06	+	135	83	27	12	9	4
5	rs2367725	1	*ST3GAL3*	0.0274	+	9.31E-06	+	459	317	56	37	31	18

p-values are for the inverse variance-weighted meta-analysis. Total represents the number of SNPs in time to death (time to event) analysis within 500 kilo base pair of SNPs from the time to event (time to death) analysis that are in LD with the top SNPs from the time to death (time to event) analysis in the HapMap CEU release 22 ($r^2 \geq 0.10$) and have association *p*-value < 0.05.
Key: Chr, chromosome; Effect, meta-analysis direction of effect; LD, linkage disequilibrium; SNP, single nucleotide polymorphism.
[a] Only SNPs that were nominally significant ($p < 0.05$) for both traits are shown.

Table 4
Results from the gene annotation analysis using PANTHER

Biological process	H. sapiens (reference)	Number of genes observed	Number of genes expected	−/+	*p*-value unadjusted	*p*-value adjusted [a]
Time to death:						
Biological process unclassified	11321	238	367.71	−	1.29E-20	4.00E-19
Developmental processes	2152	152	69.9	+	1.39E-19	4.32E-18
Neuronal activities	569	65	18.48	+	8.94E-18	2.77E-16
Signal transduction	3406	199	110.63	+	9.09E-17	2.82E-15
Neurogenesis	587	64	19.07	+	1.43E-16	2.84E-14
Ectoderm development	692	68	22.48	+	2.33E-15	3.38E-13
Cell adhesion	622	57	20.2	+	7.00E-12	2.17E-10
Time to event:						
Developmental processes	2152	115	57.46	+	1.02E-12	3.16E-11
Biological process unclassified	11321	214	302.27	−	2.93E-12	9.08E-11
Neuronal activities	569	47	15.19	+	2.28E-11	7.08E-10

Candidate genes (genes observed) were in the neighborhood of single nucleotide polymorphisms (SNPs) associated with p value $< 1 \times 10^{-3}$. For time to death 862 candidate genes were identified; 826 could be matched to the Protein ANalysis THrough Evolutionary Relationships (PANTHER) gene list. For time to event 704 candidate genes were identified; 679 could be matched to the PANTHER gene list. Extended lists of PANTHER pathways, biological processes, and molecular functions are listed in the Supplementary Tables (S13, S14).

[a] Bonferroni correction multiplying the single-test *p*-value by the number of independent tests to obtain an expected error rate.

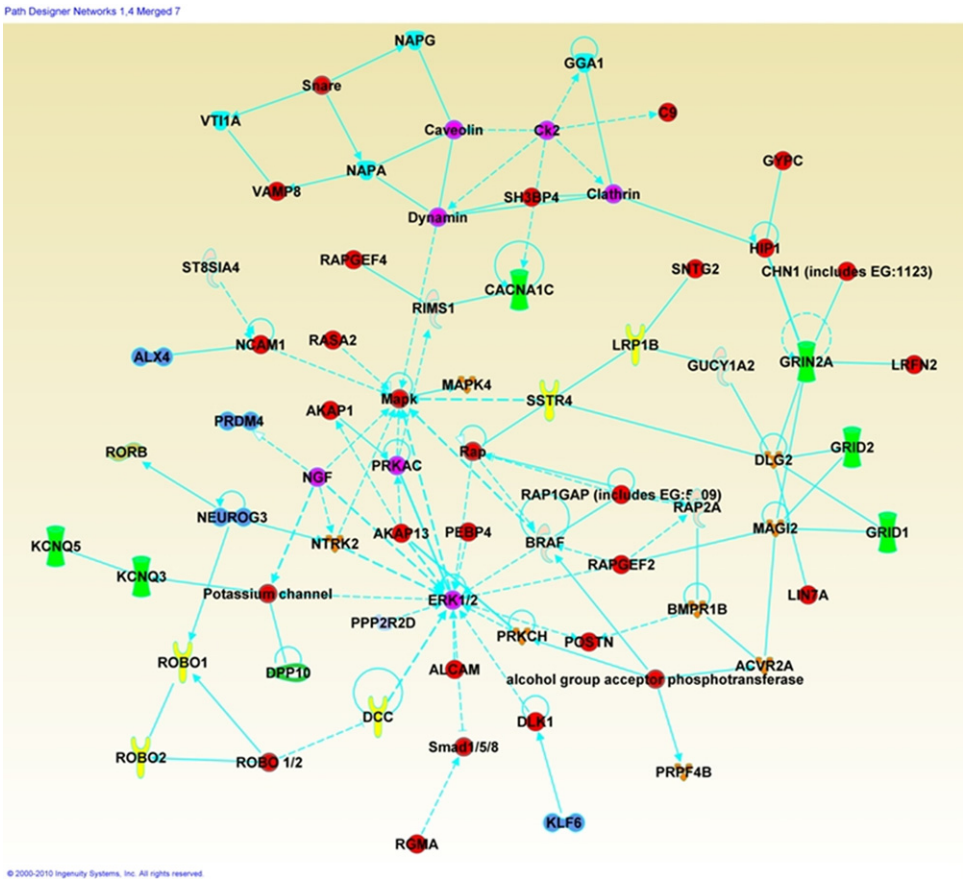


Fig. 2. Network describing neuronal activities related to time to death. Pathway analysis of genes (single nucleotide polymorphisms; SNPs) associated with time to death. Genes are represented as nodes; edges indicate known interactions (solid lines depict direct and hatched lines depict indirect interaction). Human gene functions are color-coded as follows: red = unknown, yellow = transmembrane receptor and G protein coupled receptor, magenta (pink-purple) = group/complex/other, bright green = ion channel, hunter green (dark green) = peptidase, navy blue = transcription regulator, light blue = transporter, beige = enzyme, orange = kinase, light green = cytokine, light purple = phosphate, gray = translation regulator, olive green = ligand-dependent nuclear receptor.

protein, *GRID2* (glutamate receptor, ionotropic, delta 2)—a relatively new member of the family of ionotropic glutamate receptors which are the predominant excitatory neurotransmitter receptors in the mammalian brain, and have a role in neuronal apoptotic death, and *RIMS1* (regulating synaptic membrane exocytosis 1), which regulates synaptic vesicle exocytosis and may be part of the protein scaffold of the cell.

Among the genes that were highlighted through the IPA analysis in the analysis of time to event was *MYC* (v-myc myelocytomatosis viral oncogene homolog)—a multifunctional, nuclear phosphoprotein that plays a role in cell cycle progression, apoptosis, and cellular transformation. *MYC* functions as a transcription factor that regulates transcription of specific target genes. Second in line were *E2F1* (E2F transcription factor 1), *EGFR* (epidermal growth factor receptor), and *CEBPA* (CCAAT/enhancer binding protein [C/EBP], alpha). *EF21*, a transcription factor, plays a crucial role in the control of cell cycle and action of tumor suppressor proteins can mediate both cell proliferation and p53-dependent/independent apoptosis. *EGFR* is a transmembrane glycoprotein that serves as a receptor for members of the epidermal growth factor family and supports cell proliferation. *CEBP*-Alpha, a bZIP transcription factor, can bind as a homodimer to certain promoters and enhancers. *CEBPA* also forms heterodimers with the related proteins CEBP-beta and CEBP-gamma and modulates the expression of leptin, interacts with *CDK2* and *CDK4*, and thereby inhibits these kinases and causes growth arrest in cultured cells.

4. Discussion

In our analyses of over 25,000 individuals of 55 years and older followed for an average of 11 years, we did not identify genome-wide significant associations for all-cause mortality and survival free of major diseases. However, both traits highlighted loci with suggestive significance that were in the neighborhood of genes related to neural regulation. In addition, our pathway and network analyses identified an enrichment of genes associated with cellular and neural development and function, and cell communication that may contribute to variation in human aging. Brain development might be responsible for the creation of redundancy in brain circuitry, which is associated with functional reserve and resiliency. Brain function regulates most of the compensatory strategy supporting maintenance of homeostatic equilibrium. Both of these processes are essential to healthy aging and longevity.

Several explanations are possible for the lack of genome-wide significant findings. First, mortality is arguably 1 of the most complex phenotypes, and several trajectories toward extreme old age have been identified (Evert et al., 2003). Multiple genes could mediate the aging process but would have their effects through numerous different pathophysiological processes and diseases that act as intermediate factors on the pathway to death (de Magalhaes et al., 2010). Therefore, any common variation in genes associated with aging probably has a small effect.

Second, the largely negative findings of this and other studies contrast with the intriguing animal studies of longevity. Very large effects of single genes on lifespan have indeed been observed in laboratory animals, but humans often have several homologues of these genes which might significantly differ in function or compensate for mutated genes through redundant mechanisms (Kuningas et al., 2008). This could explain why our top findings did not include genes in these pathways found in animal models. Animal models also represent genetically homogenous populations and are exposed to controlled environmental influences. The lack of replication of animal model findings in humans suggests that the use of knockout animals may not provide the optimal approach to understanding the variation in survival in humans as interactions with environmental factors may obscure the associations and prevent the identification of loci in humans.

Third, our study is based on common genetic variants and therefore we cannot exclude effects due to low frequency and rare variants ($< 5\%$) or due to the presence of structural variation, such as copy number polymorphisms. Our discovery set may lack the power to identify all the relevant loci, even though we had sufficient power to detect common SNPs (minor allele frequency = 5% or more) with a relative risk of 1.10 (Supplementary Table S2).

Last, we cannot exclude that phenotypic heterogeneity influenced our findings. While all cohorts had prospectively collected information on major health events and diagnoses, heterogeneity in the methods of assessment and classification might have limited the ability to identify true effects.

Complex diseases may result from the effects of a large number of low frequency variants, with substantial allelic heterogeneity at disease-causing loci (Pritchard, 2001; Pritchard and Cox, 2002; Swarbrick and Vaisse, 2003). Theoretical modeling that incorporates mutation, random genetic drift, and purifying selection suggests that many of the variants that affect complex traits may be in the 1%–5% frequency range (Pritchard, 2001). Indeed, sequencing of candidate genes in an attempt to capture such low frequency variants, has led to the identification of rare variants with modest effects on body mass index (Ahituv et al., 2007; Challis et al., 2002; Cone, 2000), triglyceride levels (Romeo et al., 2007), high-density lipoprotein (HDL; Cohen et al., 2004; Romeo et al., 2007) and low-density lipoprotein (LDL) cholesterol levels (Cohen et al., 2005, 2006; Kotowski et al., 2006).

It is impossible to determine the functional variant of a gene by GWAS. Moreover, we cannot conclude from the location of an SNP that this variation is involved in the expression of the closest gene. However, our top results suggested a possible role of genes involved in neurological

processes in human longevity and aging. Ten of the 22 suggestive associations identified in our analyses are in or near genes that are highly expressed in the brain (*HECW2* [Rotin and Kumar, 2009], *HIP1* [Blanpied et al., 2003], *BIN2*, *GRIA1*), were previously related to the regulation of neuronal excitability and plasticity (*KCNQ4* [Van Eyken et al., 2006], *LMO4* [Joshi et al., 2009; Leuba et al., 2004], *GRIA1*), and the maintenance of neural circuitry and synaptic plasticity (*NETO1*), or are associated with neurological diseases such as Alzheimer's disease (*LMO4* [Leuba et al., 2004], *BIN2*, GRIA1, GRIN2B), and amyotrophic lateral sclerosis (*GRIN2B*). In addition, 6 of the 22 SNPs were in close proximity to genes associated with other phenotypes of aging such as autophagy (*ATG4C* [Kenyon, 2010]), cancer (*ATG4C* [Maiuri et al., 2009], *HIP1* [Bradley et al., 2007], *HECW2* [Rotin and Kumar, 2009], *VWA5A* [Zhou et al., 2009], *MECOM*), and mitochondrial depletion syndrome (*SUCLA2*). Notably, *BIN2*, *ATG4C*, *KCNQ4*, *MECOM*, and *SUCLA2* showed associations with both traits in our study.

Using the expression quantitative trait loci (eQTL) browser (eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/) we detected eQTL associated with *HIP1*, *COL5A1*, *LOC340156*, and *SMARCA2* in time to death only.

Interestingly, SNPs known to be associated with longevity and disease in the neighborhood of APOE (Deelen et al., 2011) or *FOXO3A* (Flachsbart et al., 2009; Willcox et al., 2008) only reached nominal significance (results not shown). These genes were originally identified in studies of centenarians; it is possible that our study of cohorts comprised of individuals from the general populations did not have sufficient statistical power to identify these genes with certainty. (Tan et al., 2008).

While meta-analysis of GWAS has the power to detect small changes of allele frequencies between groups with the analyzed trait, true association signals may not be revealed based on a stringent genome-wide significance threshold. This situation, although limiting false positive findings, performs poorly in identifying false negatives as they may fall below the threshold. Network analyses using a less stringent significance threshold do not amend the overall negative finding of this study. However, it is well-recognized that within the many associations that failed to attain this level of significance lie true positive associations. Network analyses can provide useful information exploring multiple gene effects and their interactions.

In fact the interpretation of most GWAS results is difficult because individual results may involve many seemingly unrelated genes. Because PANTHER and IPA are built on different conceptual approaches, database sources and different pathway classifications, they can be seen as complementary approaches. Our pathway and network analyses highlighted neuronal activities and organism developmental processes as major biological processes involved aging. In addition, it highlighted Wnt signaling and showed that those genes that were involved in most pathways indeed had substantial effects within the analyzed trait. *NTRK2* (Rico et al., 2002), *NCAM1* (Rutishauser et al., 1988), *GRID2* (Hirai et al., 2003), and *RIMS1* (Johnson et al., 2003; Schoch et al., 2002) are associated with neuronal development and disease pathways that were highlighted in the analysis of time to death. *MYC* (Cole, 1986; Goga et al., 2007), *E2F1* (Nevins, 2001), *EGFR* (Wang et al., 2004), and *CEBPA* (Ménard et al., 2002; Wang et al., 2001) are associated with "cancer," "cell function," and "development" pathways.

Few if any of the top hits from the GWAS were involved in common pathways of aging, typically addressed in candidate gene studies. For example, there was no specific evidence for genes involved in IGF-insulin signaling. However, this negative finding cannot be interpreted as evidence against the importance of IGF-insulin signaling, as well as other processes such as inflammation, oxidative stress, cellular damage and repair, growth hormone, and cell proliferation in aging. Moreover, it is possible that polymorphisms in related genes have an effect in the oldest old, who were represented by fewer numbers in our study population such that our study design would be underpowered to detect it. It is also conceivable that the neurological pathways identified by our analysis interact with the known candidate genes involved in aging (Bishop et al., 2010; Finch and Ruvkun, 2001). It is feasible that the traditional aging pathways are hierarchically controlled by neurons and that the brain might be the location coordinating physiological changes (Bishop et al., 2010; Finch and Ruvkun, 2001). Because neurons are particularly susceptible to damage caused by reactive oxygen species, limitations in cellular maintenance and repair might reinforce these pathways and accelerate aging (Finch and Ruvkun, 2001). An increased ability of neuronal cells to prevent or repair oxidative damage might result in beneficial hormonal signaling, otherwise deregulated with age, thus delaying the onset of age-related disease and directly regulating cognitive aging and life span (Bishop et al., 2010; Cutler and Mattson, 2006; de Magalhães and Sandberg, 2005).

In conclusion, our analysis did provide suggestive evidence that aging is under neuronal control. Unfortunately, we have no relevant tissue or expression experiment available to further underscore or validate our findings. Future investigations of changes of gene expression with age at cellular and population levels are warranted.

Disclosure statement

The authors declare that no competing interests exist.

All participants included in this analysis provided written informed consent.

Replication Samples

Whitehall II: Whitehall II has been supported by grants from the Medical Research Council; Economic and Social Research Council; British Heart Foundation; Health and Safety Executive; Department of Health; National Heart Lung and Blood Institute (HL36310), US NIH National Institute on Aging (AG13196), US NIH Agency for Health Care Policy Research (HS06516); and the John D. and Catherine T. MacArthur Foundation Research Networks on Successful Midlife Development and Socio-economic Status and Health.

English Longitudinal Study of Aging: Samples from the English Longitudinal Study of Ageing (ELSA) DNA Repository (EDNAR), received support under a grant (AG1764406S1) awarded by the National Institute on Ageing (NIA). ELSA was developed by a team of researchers based at the National Centre for Social Research, University College London and the Institute of Fiscal Studies. The data were collected by the National Centre for Social Research.

Religious Order Study: Grants P30AG10161, R01AG15819, and R01AG30146 from the National Institute on Aging, and the Translation Genomics Research Institute.

Memory and Aging Project: Grants R01AG17917 and R01AG15819 from the National Instiute on Aging, and the Translation Genomics Research Institute.

The Longevity Consortium, funded by the National Institute of Aging, grant number U19 AG023122, provided administrative resources to CHARGE investigators for this phenotype as well as scientific opportunity funds for de novo genotyping of the 5 selected SNPs in the ELSA and Whitehall II replication studies.

Acknowledgements

Rotterdam Study: The Rotterdam Study is supported by Netherlands Genomics Initiative/Netherlands Consortium for Healthy Aging (050-060-810); Netherlands Organisation for Scientific Research (NWO) (904-61-090, 904-61-193, 480-04-004, 400-05-717, SPI 56-464-1419, 175.010. 2005.011, 911-03-012 and 017.106.370); Netspar—Living longer for a good health; the Erasmus Medical Center, and Erasmus University in Rotterdam; the Netherlands Organization for Health Research and Development, the Netherlands Research Institute for Diseases in the Elderly; the Dutch Ministry of Education, Culture and Science, and the Ministry for Health, Welfare and Sports; the European Commission; and the Municipality of Rotterdam, the Netherlands.

Cardiovascular Health Study: The CHS research reported in this article was supported by contract numbers N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, grant numbers U01 HL080295 and R01 HL087652 from the National Heart, Lung, and Blood Institute, the National Institute of Aging R01 AG031890 with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal CHS investigators and institutions can be found at www.chs-nhlbi.org/pi.htm. DNA handling and genotyping was supported in part by National Center for Research Resources grant M01-RR00425 to the Cedars-Sinai General Clinical Research Center Genotyping core and National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

Framingham Heart Study: Phenotype-genotype analyses were supported by the National Institute of Aging grant number R01AG029451 (PI: JMM). The Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine were supported by the National Heart, Lung and Blood Institute's Framingham Heart Study Contract No. N01-HC-25195 and its contract with Affymetrix, Inc., for genotyping services (Contract No. N02-HL-6-4278). Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. A portion of this research was conducted using the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Dr. Kiel's effort as well as all hip fracture data from the Framingham Osteoporosis Study was supported by a grant from the National Institute of Arthritis, Musculoskeletal and Skin Diseases and the National Institute on Aging; R01 AR/AG 41398. This research was additionally supported by the following grants: AG033193, AG081220, NS17950, P30AG013846, 1R01-AG028321.

Atherosclerosis Risk in Communities Study: The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, and N01-HC-55022. The authors thank the staff and participants of the ARIC study for their important contributions.

Age, Gene/Environment Susceptibility -Reykjavik Study: The Age, Gene/Environment Susceptibility-Reykjavik Study is funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). Genotyping was conducted at the NIA IRP Laboratory of Neurogenetics.

Invecchiare nel Chianti: The InCHIANTI study baseline (1998–2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health and in

part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336).

Baltimore Longitudinal Study of Aging: The BLSA was supported in part by the Intramural Research Program of the NIH, National Institute on Aging. A portion of that support was through an R&D contract with MedStar Research Institute.

Health, Aging and Body Composition: This research is supported in part by the Intramural Research Program of the NIH, National Institute on Aging. This research was supported by NIA contracts N01AG62101, N01AG62103, N01AG62106 and NIA grant 1R03AG032498-01. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C.

Study of Health in Pomerania: SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg, West Pomerania. The University of Greifswald is a member of the "Center of Knowledge Interchange" program of the Siemens AG.

Appendix. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neurobiolaging.2011.05.026.

References

Ahituv, N., Kavaslar, N., Schackwitz, W., Ustaszewska, A., Martin, J., Hebert, S., Doelle, H., Ersoy, B., Kryukov, G., Schmidt, S., Yosef, N., Ruppin, E., Sharan, R., Vaisse, C., Sunyaev, S., Dent, R., Cohen, J., McPherson, R., Pennacchio, L.A., 2007. Medical sequencing at the extremes of human body mass. Am. J. Hum. Genet. 80, 779–791.

Anselmi, C.V., Malovini, A., Roncarati, R., Novelli, V., Villa, F., Condorelli, G., Bellazzi, R., Puca, A.A., 2009. Association of the FOXO3A locus with extreme longevity in a southern Italian centenarian study. Rejuvenation Res. 12, 95–104.

Atzmon, G., Schechter, C., Greiner, W., Davidson, D., Rennert, G., Barzilai, N., 2004. Clinical phenotype of families with longevity. J. Am. Geriatr. Soc. 52, 274–277.

Bishop, N.A., Lu, T., Yankner, B.A., 2010. Neural mechanisms of ageing and cognitive decline. Nature 464, 529–535.

Blanpied, T.A., Scott, D.B., Ehlers, M.D., 2003. Age-related regulation of dendritic endocytosis associated with altered clathrin dynamics. Neurobiol. Aging 24, 1095–1104.

Brack, A.S., Conboy, M.J., Roy, S., Lee, M., Kuo, C.J., Keller, C., Rando, T.A., 2007. Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis. Science 317, 807–810.

Bradley, S.V., Holland, E.C., Liu, G.Y., Thomas, D., Hyun, T.S., Ross, T.S., 2007. Huntingtin interacting protein 1 is a novel brain tumor marker that associates with epidermal growth factor receptor. Cancer Res. 67, 3609–3615.

Challis, B.G., Pritchard, L.E., Creemers, J.W., Delplanque, J., Keogh, J.M., Luan, J., Wareham, N.J., Yeo, G.S., Bhattacharyya, S., Froguel, P., White, A., Farooqi, I.S., O'Rahilly, S., 2002. A missense mutation disrupting a dibasic prohormone processing site in pro-opiomelanocortin (POMC) increases susceptibility to early-onset obesity through a novel molecular mechanism. Hum. Mol. Genet. 11, 1997–2004.

Christensen, K., Johnson, T.E., Vaupel, J.W., 2006. The quest for genetic determinants of human longevity: challenges and insights. Nat. Rev. Genet. 7, 436–448.

Cohen, J., Pertsemlidis, A., Kotowski, I.K., Graham, R., Garcia, C.K., Hobbs, H.H., 2005. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat. Genet. 37, 161–165.

Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R., Hobbs, H.H., 2004. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305, 869–872.

Cohen, J.C., Pertsemlidis, A., Fahmi, S., Esmail, S., Vega, G.L., Grundy, S.M., Hobbs, H.H., 2006. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. Proc. Natl. Acad. Sci. U. S. A. 103, 1810–1815.

Cole, M.D., 1986. The myc oncogene: its role in transformation and differentiation. Annu. Rev. Genet. 20, 361–384.

Cone, R.D., 2000. Haploinsufficiency of the melanocortin-4 receptor: part of a thrifty genotype? J. Clin. Invest. 106, 185–187.

Cutler, R.G., Mattson, M.P., 2006. The adversities of aging. Ageing Res. Rev. 5, 221–238.

de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S., Voight, B.F., 2008. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum. Mol. Genet. 17, R122–R128.

de Magalhaes, J.P., Budovsky, A., Lehmann, G., Costa, J., Li, Y., Fraifeld, V., Church, G.M., 2009a. The Human Ageing Genomic Resources: online databases and tools for biogerontologists. Aging Cell 8, 65–72.

de Magalhaes, J.P., Finch, C.E., Janssens, G., 2010. Next-generation sequencing in aging research: Emerging applications, problems, pitfalls and possible solutions. Ageing Res Rev 9, 315–323.

de Magalhães, J.P., Sandberg, A., 2005. Cognitive aging as an extension of brain development: a model linking learning, brain plasticity, and neurodegeneration. Mech. Ageing Dev. 126, 1026–1033.

Deelen, J., Beekman, M., Uh, H.W., Helmer, Q., Kuningas, M., Christiansen, L., Kremer, D., van de Breggen, R., Suchiman, H.E., Lakenberg, N., van den Akker, E.B., Passtoors, W.M., Tiemeier, H., van Heemst, D., de Craen, A.J., Rivadeneira, F., de Geus, E.J., Perola, M., van der Ouderaa, F.J., Gunn, D.A., Boomsma, D.I., Uitterlinden, A.G., Christensen, K., van Duijn, C.M., Heijmans, B.T., Houwing-Duistermaat, J.J., Westendorp, R.G., Slagboom, P.E., 2011. Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. Aging Cell doi:10.1111/j.1474-9726.2011.00705.x.

Evert, J., Lawler, E., Bogan, H., Perls, T., 2003. Morbidity profiles of centenarians: survivors, delayers, and escapers. J. Gerontol. A Biol. Sci. Med. Sci. 58, 232–237.

Ewbank, D.C., 2007. Differences in the association between apolipoprotein E genotype and mortality across populations. J. Gerontol. A Biol. Sci. Med. Sci. 62, 899–907.

Finch, C.E., Ruvkun, G., 2001. The genetics of aging. Annu. Rev. Genomics Hum. Genet. Genet. 2, 435–462.

Flachsbart, F., Caliebe, A., Kleindorp, R., Blanché, H., von Eller-Eberstein, H., Nikolaus, S., Schreiber, S., Nebel, A., 2009. Association of

FOXO3A variation with human longevity confirmed in German centenarians. Proc. Natl. Acad. Sci. U. S. A. 106, 2700–2705.

Gerdes, L.U., Jeune, B., Ranberg, K.A., Nybo, H., Vaupel, J.W., 2000. Estimation of apolipoprotein E genotype-specific relative mortality risks from the distribution of genotypes in centenarians and middle-aged men: apolipoprotein E gene is a "frailty gene," not a "longevity gene.Genet. Epidemiol. 19, 202–210.

Goga, A., Yang, D., Tward, A.D., Morgan, D.O., Bishop, J.M., 2007. Inhibition of CDK1 as a potential therapy for tumors over-expressing MYC. Nat. Med. 13, 820–827.

Herskind, A.M., McGue, M., Holm, N.V., Sørensen, T.I., Harvald, B., Vaupel, J.W., 1996. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870–1900. Hum. Genet. 97, 319–323.

Hirai, H., Launey, T., Mikawa, S., Torashima, T., Yanagihara, D., Kasaura, T., Miyamoto, A., Yuzaki, M., 2003. New role of delta2-glutamate receptors in AMPA receptor trafficking and cerebellar function. Nat. Neurosci. 6, 869–876.

Inestrosa, N.C., Arenas, E., 2010. Emerging roles of Wnts in the adult nervous system. Nat. Rev. Neurosci. 11, 77–86.

Johnson, S., Halford, S., Morris, A.G., Patel, R.J., Wilkie, S.E., Hardcastle, A.J., Moore, A.T., Zhang, K., Hunt, D.M., 2003. Genomic organisation and alternative splicing of human RIM1, a gene implicated in autosomal dominant cone-rod dystrophy (CORD7). Genomics 81, 304–314.

Joshi, K., Lee, S., Lee, B., Lee, J.W., Lee, S.K., 2009. LMO4 controls the balance between excitatory and inhibitory spinal V2 interneurons. Neuron 61, 839–851.

Keeble, T.R., Halford, M.M., Seaman, C., Kee, N., Macheda, M., Anderson, R.B., Stacker, S.A., Cooper, H.M., 2006. The Wnt receptor Ryk is required for Wnt5a-mediated axon guidance on the contralateral side of the corpus callosum. J. Neurosci. 26, 5840–5848.

Kenyon, C.J., 2010. The genetics of ageing. Nature 464, 504–512.

Kotowski, I.K., Pertsemlidis, A., Luke, A., Cooper, R.S., Vega, G.L., Cohen, J.C., Hobbs, H.H., 2006. A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. Am. J. Hum. Genet. 78, 410–422.

Kuningas, M., Mooijaart, S.P., van Heemst, D., Zwaan, B.J., Slagboom, P.E., Westendorp, R.G., 2008. Genes encoding longevity: from model organisms to humans. Aging Cell 7, 270–280.

Leuba, G., Vernay, A., Vu, D., Walzer, C., Belloir, B., Kraftsik, R., Bouras, C., Savioz, A., 2004. Differential expression of LMO4 protein in Alzheimer's disease. Neuropathol. Appl. Neurobiol. 30, 57–69.

Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J., Dehghan, A., Glazer, N.L., Morrison, A.C., Johnson, A.D., Aspelund, T., Aulchenko, Y., Lumley, T., Kottgen, A., Vasan, R.S., Rivadeneira, F., Eiriksdottir, G., Guo, X., Arking, D.E., Mitchell, G.F., Mattace-Raso, F.U., Smith, A.V., Taylor, K., Scharpf, R.B., Hwang, S.J., Sijbrands, E.J., Bis, J., Harris, T.B., Ganesh, S.K., O'Donnell, C.J., Hofman, A., Rotter, J.I., Coresh, J., Benjamin, E.J., Uitterlinden, A.G., Heiss, G., Fox, C.S., Witteman, J.C., Boerwinkle, E., Wang, T.J., Gudnason, V., Larson, M.G., Chakravarti, A., Psaty, B.M., van Duijn, C.M., 2009. Genome-wide association study of blood pressure and hypertension. Nat Genet 41, 677–687.

Li, Y., Wang, W.J., Cao, H., Lu, J., Wu, C., Hu, F.Y., Guo, J., Zhao, L., Yang, F., Zhang, Y.X., Li, W., Zheng, G.Y., Cui, H., Chen, X., Zhu, Z., He, H., Dong, B., Mo, X., Zeng, Y., Tian, X.L., 2009a. Genetic association of FOXO1A and FOXO3A with longevity trait in Han Chinese populations. Hum. Mol. Genet. 18, 4897–4904.

Li, Y., Willer, C., Sanna, S., Abecasis, G., 2009b. Genotype imputation. Annu. Rev. Genomics Hum. Genet. 10, 387–406.

Lunetta, K.L., D'Agostino, R.B., Sr., Karasik, D., Benjamin, E.J., Guo, C.Y., Govindaraju, R., Kiel, D.P., Kelly-Hayes, M., Massaro, J.M., Pencina, M.J., Seshadri, S., Murabito, J.M., 2007. Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study. BMC Med. Genet. 8 suppl 1, S13.

Maiuri, M.C., Tasdemir, E., Criollo, A., Morselli, E., Vicencio, J.M., Carnuccio, R., Kroemer, G., 2009. Control of autophagy by oncogenes and tumor suppressor genes. Cell Death Differ. 16, 87–93.

McGue, M., Vaupel, J.W., Holm, N., Harvald, B., 1993. Longevity is moderately heritable in a sample of Danish twins born 1870–1880. J. Gerontol. 48, B237–B244.

Ménard, C., Hein, P., Paquin, A., Savelson, A., Yang, X.M., Lederfein, D., Barnabé-Heider, F., Mir, A.A., Sterneck, E., Peterson, A.C., Johnson, P.F., Vinson, C., Miller, F.D., 2002. An essential role for a MEK-C/EBP pathway during growth factor-regulated cortical neurogenesis. Neuron 36, 597–610.

Mi, H., Guo, N., Kejariwal, A., Thomas, P.D., 2007. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. Nucleic Acids Res. 35(Database issue), D247–D252.

Nevins, J.R., 2001. The Rb/E2F pathway and cancer. Hum. Mol. Genet. 10, 699–703.

Newman, A.B., Walter, S., Lunetta, K.L., Garcia, M.E., Slagboom, P.E., Christensen, K., Arnold, A.M., Aspelund, T., Aulchenko, Y.S., Benjamin, E.J., Christiansen, L., D'Agostino, R.B., Sr., Fitzpatrick, A.L., Franceschini, N., Glazer, N.L., Gudnason, V., Hofman, A., Kaplan, R., Karasik, D., Kelly-Hayes, M., Kiel, D.P., Launer, L.J., Marciante, K.D., Massaro, J.M., Miljkovic, I., Nalls, M.A., Hernandez, D., Psaty, B.M., Rivadeneira, F., Rotter, J., Seshadri, S., Smith, A.V., Taylor, K.D., Tiemeier, H., Uh, H.W., Uitterlinden, A.G., Vaupel, J.W., Walston, J., Westendorp, R.G., Harris, T.B., Lumley, T., van Duijn, C.M., Murabito, J.M., 2010. A meta-analysis of four genome-wide association studies of survival to age 90 years or older: The Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium. J. Gerontol. A Biol. Sci. Med. Sci. 65, 478–487.

Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M.D., Bochud, M., Coin, L., Najjar, S.S., Zhao, J.H., Heath, S.C., Eyheramendy, S., Papadakis, K., Voight, B.F., Scott, L.J., Zhang, F., Farrall, M., Tanaka, T., Wallace, C., Chambers, J.C., Khaw, K.T., Nilsson, P., van der Harst, P., Polidoro, S., Grobbee, D.E., Onland-Moret, N.C., Bots, M.L., Wain, L.V., Elliott, K.S., Teumer, A., Luan, J., Lucas, G., Kuusisto, J., Burton, P.R., Hadley, D., McArdle, W.L., Brown, M., Dominiczak, A., Newhouse, S.J., Samani, N.J., Webster, J., Zeggini, E., Beckmann, J.S., Bergmann, S., Lim, N., Song, K., Vollenweider, P., Waeber, G., Waterworth, D.M., Yuan, X., Groop, L., Orho-Melander, M., Allione, A., Di Gregorio, A., Guarrera, S., Panico, S., Ricceri, F., Romanazzi, V., Sacerdote, C., Vineis, P., Barroso, I., Sandhu, M.S., Luben, R.N., Crawford, G.J., Jousilahti, P., Perola, M., Boehnke, M., Bonnycastle, L.L., Collins, F.S., Jackson, A.U., Mohlke, K.L., Stringham, H.M., Valle, T.T., Willer, C.J., Bergman, R.N., Morken, M.A., Doring, A., Gieger, C., Illig, T., Meitinger, T., Org, E., Pfeufer, A., Wichmann, H.E., Kathiresan, S., Marrugat, J., O'Donnell, C.J., Schwartz, S.M., Siscovick, D.S., Subirana, I., Freimer, N.B., Hartikainen, A.L., McCarthy, M.I., O'Reilly, P.F., Peltonen, L., Pouta, A., de Jong, P.E., Snieder, H., van Gilst, W.H., Clarke, R., Goel, A., Hamsten, A., Peden, J.F., Seedorf, U., Syvanen, A.C., Tognoni, G., Lakatta, E.G., Sanna, S., Scheet, P., Schlessinger, D., Scuteri, A., Dorr, M., Ernst, F., Felix, S.B., Homuth, G., Lorbeer, R., Reffelmann, T., Rettig, R., Volker, U., Galan, P., Gut, I.G., Hercberg, S., Lathrop, G.M., Zelenika, D., Deloukas, P., Soranzo, N., Williams, F.M., Zhai, G., Salomaa, V., Laakso, M., Elosua, R., Forouhi, N.G., Volzke, H., Uiterwaal, C.S., van der Schouw, Y.T., Numans, M.E., Matullo, G., Navis, G., Berglund, G., Bingham, S.A., Kooner, J.S., Connell, J.M., Bandinelli, S., Ferrucci, L., Watkins, H., Spector, T.D., Tuomilehto, J., Altshuler, D., Strachan, D.P., Laan, M., Meneton, P., Wareham, N.J., Uda, M., Jarvelin, M.R., Mooser, V., Melander, O., Loos, R.J., Elliott, P., Abecasis, G.R., Caulfield, M., Munroe, P.B., 2009. Genome-wide association study identifies eight loci associated with blood pressure. Nat. Genet. 41, 666–676.

Pawlikowska, L., Hu, D., Huntsman, S., Sung, A., Chu, C., Chen, J., Joyner, A.H., Schork, N.J., Hsueh, W.C., Reiner, A.P., Psaty, B.M.,

Atzmon, G., Barzilai, N., Cummings, S.R., Browner, W.S., Kwok, P.Y., Ziv, E., Study of Osteoporotic Fractures, 2009. Association of common genetic variation in the insulin/IGF1 signaling pathway with human longevity. Aging Cell 8, 460–472.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909.

Pritchard, J.K., 2001. Are rare variants responsible for susceptibility to complex diseases? Am. J. Hum. Genet. 69, 124–137.

Pritchard, J.K., Cox, N.J., 2002. The allelic architecture of human disease genes: common disease-common variant. . .or not? Hum. Mol. Genet. 11, 2417–2423.

Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B., Witteman, J.C., Boerwinkle, E., CHARGE Consortium, 2009. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ. Cardiovasc. Genet. 2,73–80.

Reed, T., Dick, D.M., 2003. Heritability and validity of healthy physical aging (wellness) in elderly male twins. Twin Res. 6, 227–234.

Rico, B., Xu, B., Reichardt, L.F., 2002. TrkB receptor signaling is required for establishment of GABAergic synapses in the cerebellum. Nat. Neurosci. 5, 225–233.

Romeo, S., Pennacchio, L.A., Fu, Y., Boerwinkle, E., Tybjaerg-Hansen, A., Hobbs, H.H., Cohen, J.C., 2007. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. Nat. Genet. 39, 513–516.

Rotin, D., Kumar, S., 2009. Physiological functions of the HECT family of ubiquitin ligases. Nat. Rev. Mol. Cell Biol. 10, 398–409.

Rutishauser, U., Acheson, A., Hall, A.K., Mann, D.M., Sunshine, J., 1988. The neural cell adhesion molecule (NCAM) as a regulator of cell-cell interactions. Science 240, 53–57.

Schoch, S., Castillo, P.E., Jo, T., Mukherjee, K., Geppert, M., Wang, Y., Schmitz, F., Malenka, R.C., Südhof, T.C., 2002. RIM1alpha forms a protein scaffold for regulating neurotransmitter release at the active zone. Nature 415, 321–326.

Swarbrick, M.M., Vaisse, C., 2003. Emerging trends in the search for genetic variants predisposing to human obesity. Curr. Opin. Clin. Nutr. Metab. Care 6, 369–375.

Tan, Q., Zhao, J.H., Zhang, D., Kruse, T.A., Christensen, K., 2008. Power for genetic association study of human longevity using the case-control design. Am. J. Epidemiol. 168, 890–896.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., Narechania, A., 2003. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 13, 2129–2141.

Thomas, P.D., Kejariwal, A., Guo, N., Mi, H., Campbell, M.J., Muruganujan, A., Lazareva-Ulitsky, B., 2006. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. Nucleic Acids Res. 34(Web Server issue), W645–650.

Ulloa, F., Martí, E., 2010. Wnt won the war: antagonistic role of Wnt over Shh controls dorso-ventral patterning of the vertebrate neural tube. Dev. Dyn. 239, 69–76.

Van Eyken, E., Van Laer, L., Fransen, E., Topsakal, V., Lemkens, N., Laureys, W., Nelissen, N., Vandevelde, A., Wienker, T., Van De Heyning, P., Van Camp, G., 2006. KCNQ4: a gene for age-related hearing impairment? Hum. Mutat. 27, 1007–1016.

van Rijn, M.J., Schut, A.F., Aulchenko, Y.S., Deinum, J., Sayed-Tabatabaei, F.A., Yazdanpanah, M., Isaacs, A., Axenovich, T.I., Zorkoltseva, I.V., Zillikens, M.C., Pols, H.A., Witteman, J.C., Oostra, B.A., van Duijn, C.M., 2007. Heritability of blood pressure traits and the genetic contribution to blood pressure variance explained by four blood-pressure-related genes. J. Hypertens. 25, 565–570.

Vaupel, J.W., 1997. The remarkable improvements in survival at older ages. Philos. Trans. R. Soc. Lond. B Biol. Sci. 352, 1799–1804.

Vaupel, J.W., 2010. Biodemography of human ageing. Nature 464, 536–542.

Vellai, T., Takacs-Vellai, K., Zhang, Y., Kovacs, A.L., Orosz, L., Muller, F., 2003. Genetics: influence of TOR kinase on lifespan in C. elegans. Nature 426, 620.

Vijg, J., Suh, Y., 2005. Genetics of longevity and aging. Annu. Rev. Med. 56, 193–212.

Wang, H., Iakova, P., Wilde, M., Welm, A., Goode, T., Roesler, W.J., Timchenko, N.A., 2001. C/EBPalpha arrests cell proliferation through direct inhibition of Cdk2 and Cdk4. Mol. Cell 8, 817–828.

Wang, K., Yamamoto, H., Chin, J.R., Werb, Z., Vu, T.H., 2004. Epidermal growth factor receptor-deficient mice have delayed primary endochondral ossification because of defective osteoclast recruitment. J. Biol. Chem. 279, 53848–53856.

Willcox, B.J., Donlon, T.A., He, Q., Chen, R., Grove, J.S., Yano, K., Masaki, K.H., Willcox, D.C., Rodriguez, B., Curb, J.D., 2008. FOXO3A genotype is strongly associated with human longevity. Proc. Natl. Acad. Sci. U. S. A. 105, 13987–13992.

Zhou, Y.Q., Chen, S.L., Ju, J.Y., Shen, L., Liu, Y., Zhen, S., Lv, N., He, Z.G., Zhu, L.P., Lv, N., He, Z.G., Zhu, L.P, 2009. Tumor suppressor function of BCSC-1 in nasopharyngeal carcinoma. Cancer Sci. 100, 1817–1822.

Q Manage Health Care
Vol. 20, No. 3, pp. 234–245

"Pay-for-performance" as a Quality Improvement Tool: Perceptions and Policy Recommendations of Physicians and Program Leaders

J. Frank Wharam, MB, BCh, BAO, MPH; Matthew B. Frank, MDiv; Ann-Marie Rosland, MD, MS; Michael K. Paasche-Orlow, MD, MPH; Neil J. Farber, MD; Christine Sinsky, MD; Lisa Rucker, MD; Kimberly J. Rask, MD, PhD; Michael J. Barry, MD; M. Kathleen Figaro, MD, MS

Background: *Although pay-for-performance (P4P) compensation is widespread, questions have arisen about its efficacy in improving health care quality and consequences for vulnerable patients.* ***Objective:*** *To assess perceptions of general internists and P4P program leaders regarding how to implement fair and effective P4P.* ***Methods:*** *Qualitative investigation using in-depth interviews with P4P program leaders and focus groups with general internists.* ***Results:*** *Internists emphasized a gradual and cautious approach to P4P implementation. They strongly recommended improving P4P measure validity and had detailed suggestions regarding how. Program leaders saw a need to implement perhaps imperfect programs but with continual improvement. Both groups advocated protecting vulnerable populations and made overlapping recommendations: improving measure validity; adjusting for patient characteristics; measuring improvements in quality (vs cutpoints); and providing incentives to physicians of vulnerable populations. Internists tended to favor explicit protections, while program leaders felt that P4P might inherently protect vulnerable patients by improving overall quality.* ***Discussion:*** *Internists favored gradual P4P implementation, while P4P leaders saw an immediate need for implementation with iterative improvement. Both groups recommended specific measures to protect vulnerable populations such as*

Author Affiliations: *Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts (Dr Wharam); Program in Health Policy, Harvard University, Cambridge, Massachusetts (Mr Frank); Department of Internal Medicine, University of Michigan Health System, Ann Arbor (Dr Rosland); Department of Medicine, Boston University School of Medicine, Boston, Massachusetts (Dr Paasche-Orlow); University of California, San Diego (Dr Farber); Medical Associates Clinic and Health Plan, Dubuque, Iowa (Dr Sinsky); Albert Einstein College of Medicine, Bronx, New York (Dr Rucker); Department of Health Policy and Management, Rollins School of Public Health, Emory University, Atlanta, Georgia (Dr Rask); Medical Practices Evaluation Center, General Medicine Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts (Dr Barry); and Department of General Internal Medicine, Vanderbilt University School of Medicine, Nashville Tennessee (Dr Figaro).*

Correspondence: *J. Frank Wharam, MB, BCh, BAO, MPH, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, 133 Brookline Ave, 6th Floor Boston, MA 02114 (jwharam@partners.org).*

The authors are indebted to their focus group facilitators: Alexia Torke of the University of Chicago; Tom Staiger and Eliza Sutton of the University of Washington; and Louise Walter of the University of California, San Francisco. They are similarly grateful to Audrey Mertens Tedeman of the Massachusetts General Hospital, Lynne Caples of Vanderbilt University, and Christina Kara of Harvard Pilgrim Health Care for assistance with transcription of focus groups.

There were no sources of funding for this work and no financial support was provided to authors or contributors.

DOI: 10.1097/QMH.0b013e318222c398

improving measure validity, assessing improvements in quality, and providing special incentives to physicians of vulnerable populations.

Key words: *ethics, health policy, pay for performance, quality improvement, vulnerable patients*

Lowering health care costs and improving quality are the top priorities in the United States.[1] To motivate change, payers increasingly seek to pay physicians based on "quality index" measures that account for patient outcomes or evidence-based management.[2-3] This trend is likely to accelerate; the recent national health insurance reform legislation substantially expands Medicare "pay-for-performance" (P4P).[4]

Translating the idea of P4P into an effective[5-8] and ethical[9-10] quality improvement system will prove challenging for various reasons. For example, judging physician "quality" can be difficult in clinical scenarios that have no evidence basis or in those necessitating a balance of physician judgment and patient preferences. P4P might induce physicians to refuse care for patients whom they deem less likely to help them achieve high performance benchmarks.

Previous qualitative studies have largely focused on the current impact of P4P rather than on recommended policy changes and have examined viewpoints of key stakeholders in isolation. This study assesses and compares policy perceptions of administrators and physicians regarding how to improve the fairness and effectiveness of P4P programs. Primary care physicians face the greatest intensity of measurement under P4P[11] and are well positioned to understand how financial incentives interact with complex patient care systems. P4P program leaders have detailed knowledge of P4P designs and intended outcomes.

METHODS

Study data

Between September 2006 and October 2007, we conducted an investigation of policy and ethical issues in P4P compensation by using qualitative techniques. We evaluated 2 key stakeholder groups: leaders of major P4P programs and academic general internists.

We recruited P4P program leaders by using a "snowball" sampling technique,[12] beginning with a knowledgeable index informant and then

iteratively soliciting recommendations for others to interview. P4P program leaders were from the Centers for Medicare and Medicaid Services, Bridges to Excellence, the Agency for Healthcare Research and Quality, the LeapFrog Group, United Healthcare, the Veterans' Administration, the Center for Health Care Strategies, Integrated Healthcare Association, and the Health Information and Management Systems Society. We conducted one-on-one semistructured telephonic interviews with these informants and transcribed interviews by hand.

General internists came from 6 regions of the United States (Midwest, Mid-Atlantic, Southern, New England, Northwest, and California) and, except for the California focus group, all were attendees of regional Society of General Internal Medicine meetings. Members of the Society of General Internal Medicine are mostly practicing general internists who also conduct academic research or teach. We recruited general internists through notifications in a monthly Society of General Internal Medicine e-mail news bulletin and in the program accompanying meetings. We conducted focus groups with these participants, audiorecorded discussions, and transcribed them verbatim.

Participation in both the focus groups and P4P program leader interviews was confidential and voluntary, and we notified all participants before the discussion that aggregated material would be used for research purposes. The institutional review board of Harvard Pilgrim Health Care approved the study.

We constructed open-ended questions to generate discussion of important policy and ethical aspects of P4P on the basis of a pilot discussion with general internists at a national Society of General Internal Medicine meeting. Before focus group sessions, facilitators presented a standardized overview presentation of P4P to orient all attendees to essential P4P components. We then conducted discussions based on a topic guide. We included questions about the theoretical foundation of P4P, its benefits and ethical drawbacks, and recommendations to create fair and effective P4P systems and focused on the impact for vulnerable populations. We assessed similar themes with key informants, again using a topic guide. Focus group facilitators and the one-on-one interviewer (M.K.F.) pursued relevant and important themes based on participants' comments as well as a set of predeveloped subthemes ("probes") that the study team had identified as important during the pilot focus group.

Our primary question to internists about policy recommendations was, "How could P4P be implemented in a way that would improve the care you provide your patients without unintended adverse effects?" Probes specifically asked about recommendations for vulnerable populations, the value of articulating the goals of patients and physicians in P4P systems, and any other ideas that could improve the fairness of P4P to patients, providers, and payers. The questions for P4P program leaders were (1) "How could pay-for-performance be implemented in a way that would improve patient care while minimizing unintended ethical consequences?" (2) "How do you think specific provisions could be made to improve the outcomes of vulnerable patients using pay-for-performance?" and (3) "How do you think pay-for-performance could be made a fairer and more ethical system for all patients/ providers/ payers?" Respondents also made policy recommendations when discussing other topic questions such as when asked about the benefits and drawbacks of P4P.

One-on-one interviews lasted approximately 30 to 45 minutes and focus groups were generally 60 to 90 minutes. Study team members (J.F.W., M.B., N.F., C.S., and M.K.F.) and volunteer Society of General Internal Medicine members facilitated all discussions. The study team conducted a coding validation exercise on an initial transcript. A team of 6 investigators (J.F.W., M.P.-O., A.-M.R., M.B., N.F., and M.K.F.) then applied thematic analysis to code and categorize transcripts by using standard qualitative techniques. We also identified similar and contrasting themes between focus group participants and the one-on-one interviewees. The principal investigator (J.F.W) validated or coded all transcripts. A team of 3 investigators (J.F.W., A.-M.R., and M.B.F.) summarized the coded transcripts and validated these summaries by using MaxQDA software, VERBI GmbH Marburg, Germany. We

employed grounded theory approach to develop a comprehensive interpretation of a data set that included a large frequency of individuals and data points for each group.[13,14]

RESULTS

P4P program informants were high-level leaders of 9 major insurance and coalition P4P programs. At 1 organization, 2 leaders represented the P4P program for our interview. Seven of these leaders were men and 3 were women. Seventy-six individuals participated in 6 focus groups (Table 1). Most focus group members were practicing physicians, and approximately 57% served vulnerable patient populations. Twenty-nine percent had participated in a clinical practice P4P program, and 25% reported nonclinical experience with P4P, for example, contributing to the design of a P4P program at a veteran's administration hospital.

Coding of policy recommendations revealed 5 major themes (Table 2): (1) Research, evaluation, or careful implementation of P4P is needed; (2) Vulnerable populations in P4P systems or their physicians should be protected; (3) Health care systems and infrastructure changes are needed to make P4P fair and an accurate gauge of quality; (4) The validity of P4P quality measures should be improved; and (5) Physician involvement is crucial for optimal P4P development.

Table 1

CHARACTERISTICS OF FOCUS GROUPS PARTICIPANTS

	n =76
Region; n (%)	
California	7 (9.2)
Midwest	13 (17.1)
Northwest	12 (15.8)
Southern	23 (30.3)
Mid-Atlantic	7 (9.2)
Northeast	14 (18.4)
Age, y; n (%)	
25-35	15 (19.7)
36-45	39 (51.3)
46-55	16 (21.1)
56-65	5 (6.6)
Female; n (%)	40 (52.6)
Physicians; n (%)	73 (96.0)
Years practicing; n (%)	
0-10	26 (34.2)
11-20	31 (40.8)
21-30	13 (17.1)
31-40	1 (1.3)
Experience with P4P as clinician; n (%)	21 (29.2)
Nonclinician P4P experience ; n (%)	25 (33.8)
Practice populations served; n (%)	
Urban	52 (72.2)
Rural	8 (11.1)
Suburban	22 (30.6)
Underserved	41 (56.9)
Low income	41 (56.9)
Hours per week clinical; n (%)	
0-10	24 (31.6)
11-20	19 (25.0)
21-30	9 (11.8)
31-40	6 (7.9)
41-50	4 (5.3)
51-60	6 (7.9)

Research, evaluation, or careful implementation of P4P is needed

Internist group

All general internist groups emphasized either that research on P4P is urgently needed or that key questions about P4P must be answered before its widespread implementation. In particular, internists cited a need to determine whether P4P is truly efficacious in improving quality and to understand or minimize unintended consequences. Multiple participants stated or implied that P4P should not be widely implemented until proven safe and effective, for example:

> [I]f you're going to implement all these performance systems, I think there should be some type of evaluation on what are they actually measuring, what are some of the impacts. (California focus group participant)

Table 2

COMMON POLICY RECOMMENDATIONS OF INTERNISTS AND P4P PROGRAM LEADERS

	Common or Notable Subthemes	
Common Policy Recommendations	**General Internists**	**P4P Program Leaders**
Research, evaluation, or careful implementation of P4P is needed	How is health care quality defined?	P4P organizations should self-monitor or use published literature to determine the impact of P4P
	What are the causes of poor health care quality?	Ongoing research is needed about effective and ineffective methods of implementing P4P
	Is P4P truly effective in improving quality?	Implement P4P carefully in an evolutionary manner, learning from successes and failures
	What is the impact of patient registries/ dashboards on quality?	
	What is the impact of P4P on vulnerable populations?	
	Implement P4P very carefully or only after proven effective	
Vulnerable populations in P4P systems or their physicians should be protected	Prioritize protecting vulnerable populations and their physicians	Adjust for patient characteristics
	Use valid quality measures	Measure disparities or develop measures relevant to vulnerable populations
	Adjust for patient characteristics	Measure improvements or provide incentives to physicians caring for vulnerable populations
	Provide incentives to physicians caring for vulnerable populations	Minimize the size of bonus payments to physicians
Health care systems and infrastructure changes needed to make P4P fair and an accurate gauge of quality	Improving systems and infrastructure might be adequate for quality improvement	Information technology for monitoring quality, such as electronic medical records, is needed
	P4P should incentivize improved systems and infrastructure	
The validity of quality measures should be improved[a]	Measure at the group/system level rather than the individual level	Expand measurement to cover more realms of care and specialties
	Adjust for different case mix between physicians or groups	Use evidence-based process measures
	Measure improvements in quality rather than absolute achievement	
Physician involvement is crucial for optimal P4P development	General internists or their specialist society should help to health care quality and develop valid quality measures	Physicians or their specialty societies should help to determine appropriate P4P measures

[a]The P4P program leaders also often made recommendations about measures when asked how to protect vulnerable populations under P4P.

> [N]o performance measure should be put in place unless it's been pilot tested, has been evaluated for methodological [rigor]... (Midwest focus group participant)

Physicians expressed a need to know whether the infrastructure to measure quality exists or the minimum set of tools needed to improve quality under P4P systems are in place. Several groups stated or implied that research determining the impact on vulnerable populations was important. Physicians also suggested research on nonfinancial incentives such as patient registries/dashboards (systems that monitor and display patients who fail to reach quality standards). Other suggestions included accurately defining health care quality, determining the causes of poor health care quality, improving the science of quality measurement, using pilot studies of P4P programs before widespread rollout, and conducting longer-term studies, given that a learning curve might exist after P4P implementation.

> [W]e need to start with: do we have the infrastructures to actually measure? Do the doctors who are being measured believe the information? And begin to see how our practice of medicine might change just simply by participating in measurement... (Northwest focus group participant)

> I think we have to look at why is it that the care doesn't seem to be as good as it should be. And if you don't address that problem, giving someone an extra $5,000 or penalizing them $5,000 isn't going to change a thing. (Mid-Atlantic focus group participant)

P4P program leaders

Almost all P4P program leaders mentioned that their organization was evaluating the impact of their P4P program or that they sought out published research for guidance on implementation. They were optimistic that P4P could improve outcomes for patients and that any negative consequence of their implementation could be evaluated and amended going forward. However, suggested evaluations often did not mention ethically relevant outcomes such as expelling vulnerable patients from physician panels. P4P program leaders focused less on potential negative impacts of P4P on quality, emphasizing a potential positive or neutral effect of implementation. Several mentioned a need for ongoing evaluation, continual assessment of P4P effectiveness, or an evolutionary "learning" approach to implementation:

> Being appropriately iterative [is needed]; whatever unintended consequences occur, watching what happens. Consistent iteration can overcome it. (P4P program leader 3)

> [W]e are currently contracting with Mathematica to conduct a study on vulnerable patients in the clinical setting... We can minimize the down side by making the changes evolutionary; implementations piece by piece. (P4P program leader 6)

Vulnerable populations in P4P systems or their physicians should be protected

Internist group

There was strong agreement in all focus groups that vulnerable populations must be specifically protected under P4P. For example, participants mentioned that the creation of valid quality measures adjusted for demographic patient characteristics could help to protect disadvantaged populations from expulsion. Two groups mentioned that physicians could be provided incentives or higher bonuses to care for disadvantaged patients. Others felt that being rewarded for improved quality could help to protect vulnerable populations:

> So both measuring overall performance and measuring change over time, improvement over time, I feel like balances that a little bit so that... providers who care for traditionally disadvantaged populations are not going to be at a disadvantage with these performance measures. (California focus group participant)

> I think there's a potential, and I've kind of seen it happen, for people to discharge patients from their practice who are

> noncompliant or who are known for really not meeting goals, like the uninsured, or they have certain mental health issues—with people saying, "You know, I don't think you're a right fit for me. Maybe you should find another provider." And so I worry about those patients. I think what was commented before is to have more money flowing into those patients [which] would be an incentive for people to take [care] of them. (Northwest focus group participant)

P4P program leaders

When asked specifically about vulnerable populations, several P4P program leaders endorsed perceptions of "vulnerable patients" that were unformed before the interviews. Most had to operationalize what a "vulnerable patient" might be, and their perceptions of members of that group varied. All P4P program leaders endorsed the need to protect these patients and their physicians from undue bias. However, they did not perceive the need to focus on vulnerable patients, largely believing that a well-constructed quality evaluation would benefit all patients. They most frequently recommended risk adjustment to improve fairness:

> One solution is risk-adjustment; it's a way to combat racial disparities. By looking at case mixes you can adjust for higher proportion of minorities, etc. (P4P program leader 1)

> P4P is being very careful of measuring severity; case-mix adjusting of programs. Generally, initiators are not trying to create an incentive for cherry picking. (P4P program leader 3)

Five leaders implied that simply measuring disparities was the essential starting point to reducing them:

> [Y]ou will measure all patients. If there are disparities you will see it. We realize there are communities where the improvement is going to take special effort. (P4P program leader 4)

> By vulnerable patients I presume you mean the chronically mentally ill, those at the ends of life or early life, people who are not competent for whatever reasons. The main question is: are measures of interest to them being included in P4P. (P4P program leader 8)

A related idea voiced by several leaders was to either measure improvement or create financial or other incentives to improve the care of vulnerable populations.

> The dilemmas include payments and bonuses and balancing them. Seventy-five percent of docs are already doing well. If we don't reward performance and improvement, we may exacerbate the disparities you are seeing in patients in poorer neighborhoods that are without infrastructure. (P4P program leader 9)

In contrast, 1 leader did not believe that disparities existed in the population that his or her organization served or that it was the responsibility per se of payers to protect vulnerable patients.

> Well, I don't think these differences exist anyway so we are not measuring for them. Overall, I would say don't reward on subtypes of patients but on total population. If compliance rate is 10% and if the balance of patients is all Medicaid, it is up to the physician to bring up this issue to the payer. (P4P program leader 2)

Program leaders also suggested that lower bonus amounts (eg, less than 10% of total pay) might reduce the likelihood that physicians would expel vulnerable patients from their patient panels. They also perceived that process rather than outcomes measures could improve fairness and that continuing evaluation or an iterative process could help to develop methods for improving outcomes of vulnerable patients under P4P.

Health care systems and infrastructure changes needed to make P4P fair and an accurate gauge of quality

Internist group

All focus groups emphasized the importance of health systems, care teams, or infrastructure in

improving health care quality. Physicians either mentioned that P4P should encourage improved systems or that systems, rather than individual physicians, should be measured and rewarded. Some felt that infrastructure improvement was adequate to improve quality in the absence of financial incentives. Several groups mentioned specific infrastructure that could improve quality with or without P4P, such as electronic medical records, patient dashboards/registries, or added ancillary staff.

> [I]f you can have this greater alignment of the incentives and the structure that would facilitate those, that would kill all birds with one stone... And kind of going back to the system structural level actually might be the most effective way of improving care, at the same time minimizing all potential downsides. (California focus group member)

> And the analogy is that if you have 10 doctors in a group and you get $50,000 extra, well then maybe you can hire computer consultant to finally come and get the electronic record. Or you can hire a part-time person who is going to get on the phone and say "Hi, Mrs. Jones, you haven't had a Pap smear in 2 years, why don't we schedule that for you?" Using the money to do what it is intended to do. (Mid-Atlantic focus group member)

P4P program leaders

The P4P program leaders presumed a more robust health care infrastructure and system compared with internists. Leaders mostly focused on the need for improved information technology such as electronic medical records for tracking quality information, aggregating it, and then making necessary iterative changes to the P4P program.

> Health care is information; the better data goes towards better management... The lack of information sharing among different providers is likely killing patients. This is a terrible tragedy, the lack of IT. Information is not good if it can't come with you. (P4P program leader 5)

> We are moving towards outcome-based payments. We aren't there yet; we need accurate data and universal electronic medical records. (P4P program leader 7)

The validity of P4P quality measures should be improved

Internist group

General internists focused heavily on how measurement could be made more valid, and 3 common themes emerged. Almost all groups mentioned the tension between individual physician and practice-level compensation and the potential effect on quality measurements. Participants expressed concerns about individual incentives and penalties when performance is measured too narrowly or superficially; they perceived that assessment at a systems level could improve the validity of quality measures:

> [There are] a lot of deficits which could potentially be helped by having performance measures based on system-level characteristics and care processes. (California focus group participant)

> And that is where I think this whole systems thing comes into play; where, whether they give me money or not, I cannot make any changes. I just can't. I am working within this ridiculous system and it is not until they give the system an incentive that anything is going to actually change. (Mid-Atlantic focus group participant)

Typically, at least 1 participant in each focus group expressed the opinion that group level measurement was inferior because it was less effective in motivating individual physician change.

> [S]ometimes [physicians] can use the system problems as a cop out and not take personal responsibility for what we can do to change the system or for advocating for changing the system. So maybe there is something to being punished because your system is bad, because it means everybody sinks or swims together. (California focus group participant)

A second recommendation to improve measurement validity was the use of risk adjustment to

account for different case mix between physicians or groups.

> That between issues of sample size and issues of case-mix adjustment, these are things that can make the complete difference between having meaningful numbers and meaningless numbers. And if you incentivize meaningless results, then you are going to create more chaos in the system. (Mid-Atlantic focus group participant)

> I take care of an under-served population and it is a huge problem. We haven't gone to pay-for-performance—obvious issues of adverse selection. We need to adjust for case mix index in order for pay-for-performance to really work in these populations... If we don't adjust for these populations, we are going to be blocking people out of practices if this becomes a huge incentive in the wrong way. (Mid-Atlantic focus group participant)

Physicians were generally concerned that rewarding only certain high achievement levels would divert funds to groups already performing at high levels, give little incentive for improvement to groups already performing at high or low levels, and put physicians caring for vulnerable populations at a disadvantage.

> And really the perfect system is to say you've got a population of patients and you can't do exceptions and your job is to improve the status of that population rather than necessarily fitting a specific point. (New England focus group participant)

> And it seems like something that the VA is already doing, is something that I'd push strongly for—something that only can work is you have robust data systems—is to track improvements over time along with one-time measures because there you have two types of measures that have different biases. (California focus group participant)

P4P program leaders

Almost all P4P program leaders stated that quality incentives should be objective, measurable, and amenable to tracking; they believed that these same incentives would equally protect vulnerable populations. Given their belief in the validity of P4P goals, 4 P4P program leaders recommended expanding incentives to cover more aspects of care and specialties:

> Expanded measures are needed; more objective measures of a lot more things; I know physicians like yourself don't agree, but a lot more things happen than are being measured; and the more measurement, the more things that are not even being measured improve. (P4P program leader 4)

> So the question is equity. A lot of the population that need these services don't have quality measures. The quality of the performance measures is not great and not equitable across segments of the populations that we serve. (P4P program leader 8)

Four leaders perceived that either evidence-based measures or process measures led to better patient outcomes; they understood the possible ethical tension between patients' preferences that could impact outcomes and physician behavior, which drives process measures to a greater extent:

> Outcomes-based performance is a huge ethical dilemma; for instance, number-one fertility clinics don't take hard-to-treat patients since they don't want to worsen their statistics... You could not ignore process measures; they could be weighted. (P4P program leader 5)

> I think we need to focus on process that is evidence-based. Process needs to be very specific; the P4P must be structured from the provider's point of view. The measurements should be specific to, for example, cardiac surgeons, psychiatrists, internists. (P4P program leader 8)

One notable recommendation was that providers take the initiative to prove to payers that they already provide high-quality care, as the respondent had seen in 1 instance:

> There is a new trend, of providers trumping payers, trying to improve outcomes themselves. For instance, there is a small heart practice where the payer sent them claims info [asking them to change their practice]. They had clinical data, proving that it would be wrong for them to change; the payer learned from their superior clinical data and it improved its P4P program. (P4P program leader 5)

Physician involvement is crucial for optimal P4P development

Internist group

Focus group participants perceived that general internists' involvement was necessary in the development of P4P, especially in defining health care quality, physician excellence, and valid measures. Participants often lamented that P4P was a reality that must be accepted, though it was far from ideal:

> [W]e also have an ethical obligation to try to make constructive policy suggestions that will make an inevitable program better and more responsive to the needs of generalists rather than just pointing out the downside. (Mid-Atlantic focus group participant)

> So I think this is an opportunity to actually help define what quality is and... of determining a methodology. And actually taking a stand on what it is that we feel we should be evaluated on, not [leaving P4P development up to] some outside organization, private, for-profit company. (California focus group participant)

Other participants endorsed a greater need for patient involvement either in the development of incentives or in the measurement of physician quality. These same participants suggested that that corporate influence should be minimized in the development of P4P measures.

P4P program leaders

When asked about what physicians could do to help to improve P4P, almost all program leaders endorsed the involvement of physicians or their specialty societies in improving P4P. The most common theme was the perception of greater validity for quality measures that had physician input. Two leaders perceived a great need for physician involvement in P4P research and implementation and offered a warning about physician unwillingness to become active.

> Since P4P is inevitable, we appreciate physicians' concerns, but it's already here. So [you] can help do it or it will be done to you. (P4P program leader 5)

DISCUSSION

We assessed the policy recommendations of primary care physicians and P4P program leaders regarding performance incentives. Compared with P4P program leaders, general internists perceived a greater need to employ a highly cautious approach to implementation. Internists also focused more on improving the validity of P4P measures and had more detailed suggestions regarding how this could be accomplished. Both internists and P4P program leaders understood the importance of protecting vulnerable patients, although they had divergent suggestions on how to implement such safeguards. Their suggestions included improving the validity of measures, adjusting for patient characteristics, measuring improvements in quality, and providing "improvement" incentives to physicians caring for vulnerable populations. While physicians expressed a concern for fairness in implementation, program leaders might have perceived that all patients, including vulnerable populations, would benefit from P4P. Finally, both groups endorsed an urgent need for the direct involvement of physicians in creating or improving P4P systems.

Several previous studies have analyzed the attitudes of key stakeholders regarding P4P, though none have compared policy recommendations of physicians and P4P program leaders. Weinick and colleagues[15] interviewed hospital executives about

the use of P4P to reduce racial/ethnic disparities in care. Similar to our findings, stakeholders recommended using financial incentives to reward safety net hospitals for reducing disparities. Another study found that P4P program leaders recommended that performance improvement as well as absolute achievement should be rewarded and that measures of disparities should be developed.[16] Physicians in the United Kingdom raised concerns regarding the validity of performance targets for chronic kidney disease.[17] Neuman and colleagues[18] found that surgeons were skeptical about the use of quality measures at the individual level, and most felt they should not be used in P4P programs. A survey demonstrated that nonprimary care physicians support grants to improve infrastructure, such as electronic medical records.[19]

Our study has several limitations. Our findings might not be representative of primary care physicians or P4P program leaders. However, internists came from all areas of the country and had a range of demographic and patient panel characteristics. We included leaders of some of the most prominent US P4P programs whose views are important in shaping the direction of performance incentive programs. In addition, our aim was not to determine representative views but to gather recommendations that might have a beneficial impact on physician compensation policy. Our results might be biased by selection effects; for example, more ardent opponents of P4P might have attended focus groups or more highly committed P4P program leaders might have agreed to interviews. Nevertheless, our technique of interviewing informants from opposite ends of the opinion spectrum provides a counterbalance. We found that the major policy recommendations of the 2 stakeholder groups contained substantive overlap, so that more extensive sampling might not be highly revealing. Because we assessed a topic controversial to physicians and focused on vulnerable populations, social desirability bias could have influenced our findings. Physicians and P4P program leaders might have felt that it was more acceptable in the interview environment to advocate for vulnerable populations. Physicians might have been more likely to criticize P4P when surrounded by other physicians, given that studies have found that many physicians have negative attitudes about payers. The P4P program leaders could have altered their views to be less enthusiastic about P4P because they were being interviewed by a primary care physician.

Our research has important policy implications. Despite physicians' substantial reservations about P4P, we found overlap with program leaders regarding recommendations to improve health care quality or at least minimize unintended consequences of P4P. Both groups supported cautious P4P implementation, although physicians recommended a much more methodical and meticulous approach than has been realized to date. Steps recommended by general internists include ensuring appropriate infrastructure for quality improvement, defining quality in a manner acceptable to physicians, carefully developing measures that assess genuine quality, and then testing for efficacy before widespread use. Both stakeholder groups acknowledged the need for a strong emphasis on creating measures and incentives that protect vulnerable populations. Physicians and P4P program leaders also acknowledged that direct physician involvement in designing compensation systems is important, and performance incentives seem unlikely to succeed without physician acceptance. Nevertheless, physician focus groups generally had a visceral "us against them" tone, and the involvement of a minority of physicians in high-level design decisions might be unlikely to change this negative dynamic. Thus, although such high-level physician involvement is fully necessary, it is unlikely to be sufficient. Allowing physicians to choose individualized improvement options such as reporting rather than incentives or tailoring measurement to their specific panels might be much more beneficial in reducing the sense of powerlessness that physicians sometimes experience under P4P. For their part, P4P program leaders had a tone of frustration with the pace of research and physicians whom they viewed as resistant to change or only interested in protecting reimbursement. They also recognized the limitation of "top-down" tools in improving quality. Physicians and their societies should therefore

demonstrate commitment to quality improvement—not simply fair reimbursement—through their advocacy efforts.

Policy makers should strongly consider the perceptions of the stakeholder groups that we interviewed, with particular attention to harmonizing the attitudes of physicians and P4P program leaders. Further study should bring physicians and P4P program leaders together to establish detailed plans for performance incentives that will optimally improve health care quality.

REFERENCES

1. Committee on Quality of Health Care in America IoM. Crossing the quality chasm: a new health system for the 21st century, 2001. http://www.nap.edu/catalog/10027.html#toc. Accessed January 27, 2009.
2. Rosenthal MB, Fernandopulle R, Song HR, Landon B. Paying for quality: providers' incentives for quality improvement. *Health Aff (Millwood).* 2004;23(2):127-141.
3. Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. Pay for performance in commercial HMOs. *N Engl J Med.* 2006;355(18):1895-1902.
4. Jost TS. Health insurance exchanges: legal issues. *J Law Med Ethics.* 2009;37 (suppl 2):51-70.
5. Rosenthal MB, Frank RG. What is the empirical basis for paying for quality in health care? *Med Care Res Rev.* 2006;63(2):135-157.
6. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med.* 2006;145(4):265-272.
7. Fairbrother G, Hanson KL, Friedman S, Butts GC. The impact of physician bonuses, enhanced fees, and feedback on childhood immunization coverage rates. *Am J Public Health.* 1999;89(2):171-175.
8. Roski J, Jeddeloh R, An L, et al. The impact of financial incentives and a patient registry on preventive care quality: increasing provider adherence to evidence-based smoking cessation practice guidelines. *Prev Med.* 2003;36(3):291-299.
9. Fisher ES. Paying for performance—risks and recommendations. *N Engl J Med.* 2006;355(18):1845-1847.
10. Wharam JF, Paasche-Orlow MK, Farber NJ, et al. High quality care and ethical pay-for-performance: a Society of General Internal Medicine policy analysis. *J Gen Intern Med.* 2009;24(7):854-859.
11. Baker G, Carter B. *Provider Pay-for-Performance Incentive Programs: 2004 National Study Results,* 2005. http://www.medvantageinc.com/Pdf/MV_2004_P4P_National_Study_Results-Exec_Summary.pdf. Accessed December 2, 2010.
12. Rice P, Ezzy D. *Qualitative Research Methods: A Healthy Focus.* New York, NY: Oxford University Press; 1999.
13. Strauss AL, Corbin JM. Grounded Theory in Practice. Thousand Oaks: Sage Publications;1997.
14. Patton MQ. Qualitative Research and Evaluation Methods. 3ed. Thousand Oak, Calif.: Sage Publications; 2002.
15. Weinick RM, Chien AT, Rosenthal MB, Bristol SJ, Salamon J. Hospital executives' perspectives on pay-for-performance and racial/ethnic disparities in care. *Med Care Res Rev.* 2010;67(5):574-589.
16. Chien AT, Chin MH, Davis AM, Casalino LP. Pay for performance, public reporting, and racial disparities in health care: how are programs being designed? *Med Care Res Rev.* 2007;64(5)(suppl):283S-304S.
17. Crinson I, Gallagher H, Thomas N, de Lusignan S. How ready is general practice to improve quality in chronic kidney disease? A diagnostic analysis. *Br J Gen Pract.* 2010;60(575):403-409.
18. Neuman HB, Michelassi F, Turner JW, Bass BL. Surrounded by quality metrics: what do surgeons think of ACS-NSQIP? *Surgery.* 2009;145(1):27-33.
19. Murphy KM, Nash DB. Nonprimary care physicians' views on office-based quality incentive and improvement programs. *Am J Med Qual.* 2008;23(6):427-439.

ORIGINAL RESEARCH

Determinants of Mammography in Women With Intellectual Disabilities

Joanne E. Wilkinson, MD, MSc, Emily Lauer, MPH, Karen M. Freund, MD, MPH, and Amy K. Rosen, PhD

Background: **Women with intellectual disabilities have the same rate of breast cancer as other women but are less likely to undergo screening mammography. Characteristics associated with mammography for women with intellectual disabilities in the United States are unknown.**

Methods: **This study was based on a secondary data analysis of the Massachusetts Department of Developmental Services database, comparing women who had a mammogram within 2 years with women who had not on variables related to the ecological model. Bivariate analyses, logistic regression, and assessment of interactions were performed.**

Results: **The study sample's (n = 2907) mean age was 54.7 years; 58% lived in 24-hour residential settings, 52% received nursing health coordination, and more than 25% had clinical examination needs (eg, sedation). Residential setting, health coordination, and recent influenza vaccination were all associated with mammography. Having a guardian, higher level of activities of daily living needs, and examination needs (requiring sedation or limited wait time for examinations) were associated with lower rates. Interactions between health coordination and examination needs confirmed the potential of the nurse to ameliorate barriers to mammography.**

Conclusion: **Several system-level variables were significantly associated with mammography and, in some cases, seemed to ameliorate intrapersonal/behavioral barriers to mammography. Community agencies caring for intellectually disabled women have potential to impact mammography rates by using health coordination. (J Am Board Fam Med 2011;24:693–703.)**

Keywords: **Cancer Screening, Learning Disabilities**

Adults with intellectual disabilities are increasingly likely to live in the community and be cared for by community primary care practices.[1] Intellectually disabled adults are known to have health disparities,[2] especially regarding preventive care and health screening.[3,4] National efforts are underway[5,6] to identify the sources of these disparities and to improve screening/preventive services for adults with intellectual disabilities. One important disparity for women with intellectual disabilities is in breast cancer screening.[7] Although the rate of breast cancer among women with intellectual disabilities[8,9] mirrors that of the general population, they seem to have higher mortality rates from

This article was externally peer reviewed.

Submitted 15 March 2011; revised 13 June 2011; accepted 22 June 2011.

From the Departments of Family Medicine (JEW) and General Internal Medicine (KMF), the Women's Health Interdisciplinary Research Center (KMF), School of Medicine; and the Department of Health Policy and Management, School of Public Health (AKR), Boston University, Boston, MA; the Center for Developmental Disabilities Evaluation and Research, University of Massachusetts Medical School, Waltham, MA (EL); and the Veterans Affairs Healthcare System Center for Organization, Leadership, and Management Research, Boston, MA (AKR).

Funding: This project was supported by award no. K07CA 134547 from the National Cancer Institute.

Prior presentation: Previously presented in part at the Northeast regional meeting of the Family Medicine Education Consortium, October 2008, Baltimore, MD, and October 2010, Hershey, PA; the North American Primary Care Group Annual Meeting, November 2009, Montreal, Canada, and November 2010, Seattle, WA; and the American Society of Preventive Oncology Annual Meeting, March 2010, Bethesda, MD.

Conflict of interest: none declared.

Corresponding author: Joanne E. Wilkinson, Department of Family Medicine, Dowling 5, Boston University School of Medicine, 840 Harrison Avenue, Boston, MA 02118 (E-mail: Joanne.Wilkinson@bmc.org).

See Related Article on Page 628.

breast cancer (30.9 per 100,000 women with intellectual disabilities vs 24.2 per 100,000 women in the general population of women in Massachusetts since 2002).[10,11] Though these estimates are based on small sample sizes, it is a provocative preliminary finding given that intellectually disabled women are also thought to have lower screening rates compared with the general population. The Center for Disease Control and Prevention's Behavioral Risk factor Surveillance System data report 75% of US women older than 40 having a mammogram, and in Massachusetts that number is 85%.[12] Rates of mammography among women with intellectual disabilities are significantly lower internationally[13,14] (12% to 30%) and are unknown in the US. Researchers are investigating individual-, caregiver-, and system-level issues linked to this disparity.[15] Because most people with intellectual disabilities live in the community and are patients of neighborhood primary care practices, it is important to consider their unique vulnerabilities and barriers to screening when striving to deliver care in the patient-centered medical home.

An intellectual disability is defined as an intelligence quotient at least 2 standard deviations below the mean accompanied by significant difficulties in at least one area of adaptive functioning: conceptual (language); social (understanding of and ability to follow rules, gullibility, interpersonal relationships); and practical (activities of daily living, taking medications, handling money).[16] These difficulties must exist before age 18. Because the diagnosis of intellectual disabilities encompasses multiple areas of function, the ecological model[17] is most appropriate when considering theories of health behavior applicable to breast cancer screening (Table 1), particularly the intra- and interpersonal and "institutional" (referred to herein as "system-level") domains. This model is particularly well-suited to women with intellectual disabilities because they may make health care decisions with some amount of support—from family, staff, or a guardian/agency charged with their care. Therefore, the multiple environmental domains of this model provide a broader perspective for intellectually disabled women. Interactions between domains also may be relevant for understanding associations with screening mammography. The aim of this study was to determine characteristics associated with mammography, both variables and interactions between variables, related to domains of the ecological model. A secondary aim was to make preliminary recommendations based on these findings for interventions to improve screening and prevention of breast cancer in women with intellectual disabilities in the setting of the patient-centered medical home.

Methods

Database

The Massachusetts Department of Developmental Services (DDS) began collecting and tracking health information on clients with intellectual dis-

Table 1. Proposed Domains of the Ecologic Model Affecting Breast Cancer Screening for Women with Intellectual Disabilities

Concept	Concept Description	Example of Effect on Breast Cancer Screening	Related Variables of Interest
Intrapersonal	Individual factors or ideas influencing behavior	Extremely anxious about mammogram so does not have one	Psychiatric diagnoses Requires sedation or other accommodations for clinical visits Down syndrome Functional status
Interpersonal	Social supports, family, peer groups influencing behavior	Supportive guardian encourages patient to go and accompanies her	Communication status Whether guardian is assigned Whether subject is receiving other screening/preventive services
Institutional	Rules and policies that may promote or prevent behavior	Residential program provides care coordination by a nurse and encourages cancer screenings	Residential setting Day/work program Care coordination by registered nurse
Community	Social groups/organizations in the community that can be formal or informal	Advocacy organizations for adults with disabilities publishes information encouraging mammography	None in this database
Public policy	Local policies and laws to support healthy behaviors	Publications raise awareness of physicians about preventive services for adults with disabilities	None in this database

abilities within the last 10 years. We obtained data for this study from these administrative "health record" entries of this electronic client management database, which was tailored for the needs of adults with intellectual disabilities and included information about functional status and special needs related to medical care. It is important to note that these records are not the same as electronic medical records used by health care providers. The database is used to track health outcomes for clients of DDS, not to provide medical care. It was determined through prior analyses[18] that the database does not have uniform representation of all clients. The database is most generalizable to women with intellectual disabilities living in residential settings with 24-hour support (>90% of clients in this group were represented in the database). The mammography data in the DDS database were validated using matched records from a large electronic medical record database in Boston,[18] and a high correlation in data element reliability was found between the databases.

The state requires annual updates of this record by the service provider and recommends updating whenever the individual's information changes significantly. Information about each subject's service enrollment, such as state-funded residential programs, was taken directly from enrollment tables, which are updated frequently, when enrollment determines payment to the service provider.

Subjects were included in the analyses if they were women 42 to 74 years old on January 1, 2007 (to ensure that all were eligible for mammography for the entire time period), eligible for state services for at least 1 month between October 2007 and April 2009, and had complete records. To eliminate exposure time bias, subjects were included only if they had documentation for the entire time period. Reporting bias was minimized by collecting data on mammograms completed between January 1, 2007, and December 31, 2008, but entered into the database between December 31, 2008, and April 30, 2009. Women with a history of breast cancer were excluded. Also excluded were approximately 200 women who were 75 years of age or older because data currently reflect no mortality benefit from screening women in this age group.[19]

Variables

Dependent

Appropriate breast cancer screening was defined as having received a mammogram between January 1, 2007, and December 31, 2008 (yes or no). During this time period, most guidelines recommended mammography every 1 to 2 years starting at age 40.[19] (We did perform a sensitivity analysis looking only at subjects aged 50 years and older because of the revised recommendations in 2009). "Unknown" mammograms were classified as not completed because the scheduling/logistic issues involved (transporting the person to a separate test) make it likely that either the subject or her caregiver would remember the mammogram if it had occurred.

Independent

Information about independent variables was captured for each subject at baseline to 6 months before the period in which mammography screening was examined. Several variables were examined related to the intrapersonal domain of the ecological model. Age was analyzed categorically (40–49 years, 50–59 years, 60–69 years, and ≥70 years). A summary score of 0 to 4 was created for functional status based on assistance needs with four activities of daily living (ADLs): toileting, eating, personal hygiene, and ambulation. A separate variable was included for psychiatric diagnoses (one or more vs none, two or more vs none, three or more vs none; sensitivity analyses compared this classification with a classification by type of psychiatric diagnosis) and another for Down syndrome (because of possible lower rates of breast cancer).[20,21] Variables related to clinical visits were examined: need for special positioning, sedation, or limited waiting times or a tendency to be "uncooperative" during medical visits. Variables from the interpersonal domain included communication (able vs unable using any modality) and the assignment of a guardian.

Variables were examined related to the receipt of other preventive services (Papanicolaou smear, influenza vaccination after 2007, and colonoscopy/sigmoidoscopy and bone densitometry for women older than 50 years) as belonging to the "institutional," or system-level, domain. Receipt of the influenza vaccine was selected as a recent care marker in multivariate analyses because it is an easily administered annual preventive measure generally recommended for this population,[22] and its receipt likely signifies that the support staff and agencies involved with the client are pursuing preventive services for them.

Other variables related to the system-level domain were also examined, including several catego-

ries of residential setting: state-funded 24-hour support (usually provided in group home settings) and less than 24-hour support, which is a combination of shared living, subjects living independently or with family, or subjects receiving limited support at home. Health coordination by a registered nurse (RN) was also examined because many subjects receive health coordination by nurses familiar with the health needs of this population. Nurses review clients' records and make recommendations regarding medications, side effects, chronic conditions, and testing/evaluation. The RN is involved in planning for physical examinations and/or accompanies the patient to the visit, so they may influence the receipt of preventive services. Health insurance was not examined; more than 95% of the participants had Medicaid benefits.

Statistical Analyses

Bivariate analyses identified variables associated with mammography. To assess for multicollinearity, a Pearson correlation matrix was constructed between all variables considered for inclusion in multivariate regression, and tolerance/variance inflation factors were reviewed. A multivariate logistic regression model was built through stepwise reduction; the dependent variable was "recent mammography." The multivariate model was first built with univariate-level variables, using a significance criterion of 0.05 for the Wald χ^2 as the elimination threshold. This step was repeated using Akaike Information Criterion statistics[23] and yielded similar results. Interactions were tested across domains of the ecological model based on consultation of the literature and health care experts in intellectual disabilities. Two-way interactions (age as a categorical variable, care coordination by a nurse, and 24-hour supported residential setting) were tested for all variables. Finally, a three-way interaction between guardian status, nursing coordination, and being "uncooperative" or requiring limited waiting times during examinations was tested to examine how the presence of a guardian affects the interaction between these two predictors. SAS software (version 9.2, SAS Inc, Cary, NC) was used for all analyses. Several sensitivity analyses were also performed (see Table 2). For example, we examined the subset of women who did not receive the influenza vaccine; within this subset, we compared those who did and did not receive breast cancer screening.

Multivariate Model's Predictive Ability

To assess the model's sensitivity and specificity, the estimated model was applied to the dataset, and the

Table 2. Methods and Results of Sensitivity Analyses Performed

Variable of Interest	Methods Used to Perform Sensitivity Analyses	Results
ADLs	Each individual ADL score was compared to the summary score. Two different grouped levels of summary score were tested.	No significant improve in the model
Psychiatric diagnosis	Types of psychiatric diagnoses (eg, anxiety, psychosis) were compared to the number of diagnoses recorded (one, two, three or more vs none).	No significant improve in the model
Age	Categorical groupings in Table 1 were compared with continuous variable versus categorical groupings with the last category of ≥60 years.	No significant improve in the model
24-Hour residential setting	Entire analysis was re-run using only clients from 24-hour residential settings because their representation in the database was relatively complete.	All variables remained in the model except guardian and summary ADL score. Effect sizes were similar but slightly higher for all remaining variables in the model. C statistic = 0.723
Recent influenza vaccination	Characteristics of influenza vaccine were negative; women who did not receive a mammogram were analyzed and compared with women who did receive a mammogram and had influenza.	Less able to communicate (64% vs 79%) More likely to have a guardian (58% vs 46%) Less likely to have one or psychiatric diagnosis (54% vs 62%) More likely to have high ADL need (31% vs 19%) More likely to require sedation (22% vs 15%)

ADL, activity of daily living.

model's predicted mammography outcome for each subject was compared with their observed outcome.

Results

There were 2907 subjects included in the analysis. One hundred ninety-five records (6%) were excluded because of missing values. The average age of the cohort was 54.7 years (median, 53.6 years; SD, 8.2 years; range, 42.0–74.9 years).

The overall mammography rate was 53%. Table 3 shows the bivariate analyses of mammography receipt. All the variables except age show statistically significant ($P < .05$) associations with mammography. In the intrapersonal domain, all the categories reflecting higher need for support (needing special positioning, uncooperative during examinations, higher ADL need) were associated with lower odds of mammography (odds ratios [ORs] ranging from 0.69 to 0.84) except psychiatric diagnoses: subjects with a higher number of diagnoses had higher odds of receiving a mammogram. Among system-level variables, residential setting (unadjusted OR, 1.32; 95% CI, 1.14–1.53) and health coordination by an RN (OR, 1.40; 95% CI, 1.21–1.63) are most strongly associated with mammography. All the preventive care variables were strongly associated with mammography, with recent influenza vaccination being the strongest (OR, 4.38, 95% CI 3.74 to 5.12).

Table 4 shows results from the multivariate regression model. After adjusting for other variables in the model, the system-level factor most positively associated with mammography was receipt of influenza vaccination in the same time period (adjusted OR, 4.67; 95% CI, 3.84–5.66). Intrapersonal factors such as high ADL need (adjusted OR, 0.68; 95% CI, 0.55–0.84); requiring special positioning for examination (adjusted OR, 0.65; 95% CI, 0.44–0.95); and having Down Syndrome (adjusted OR, 0.63; 95% CI, 0.48–0.82) were associated with *lack* of mammography. Family history of breast cancer was positively associated with mammography (adjusted OR, 1.91; 95% CI, 1.35–2.70). Finally, two interpersonal variables showed significant associations: ability to communicate (adjusted OR, 1.44; 95% CI, 1.14–1.81) was positively associated with mammography, and assignment of a guardian (adjusted OR, 0.77; 95% CI, 0.61–0.95) was negatively associated with mammography. The C statistic for the final model was 0.728.

Two interactions illustrate the mitigating effect of system-level factors on barriers to mammography presented by intrapersonal factors. A significant interaction was noted between the subjects labeled "uncooperative" or limited waiting period and having health coordination by RN. Subjects who were labeled "uncooperative" or required a limited wait time for examinations were less likely to obtain mammography (adjusted OR, 0.79; 95% CI, 0.71–0.89) than those who were cooperative and did not require a limited wait. However, when "uncooperative" subjects also had health coordination, they did not exhibit significantly different odds of mammography (adjusted OR, 0.92; 95% CI, 0.81–1.05) compared with subjects who were considered cooperative. In addition, a significant interaction was noted between a subject's ADL score and the presence of 24-hour residential supports. Subjects with high daily assistance needs (support needs in three or more domains across four total) were less likely to receive mammograms if they received less than 24-hour residential supports (adjusted OR, 0.77; 95% CI, 0.68–0.87). In comparison, subjects with similar support needs who received 24-hour residential supports had odds of receiving a mammogram (adjusted OR, 0.88; 95% CI, 0.78–1.01) statistically similar to subjects in this setting with lower support needs. Results of the sensitivity analyses are summarized in Table 2.

We also looked at the effect of removing subjects in the 40- to 50-year-old age range from the analyses because the US Preventive Services Task Force guidelines were reissued during the course of this research project and emphasized routine screening mammography in women 50 and older. We found that when the analyses were repeated with subjects only 50 years of age and older, the overall findings were quite similar (eg, the effects of residential setting, health coordination by RN, requiring sedation for visits, ADL status) but that a few variables were not included in the final model: communication status, having a guardian, and needing special positioning for examinations (the last variable did not reach statistical significance because of the smaller sample size when women 40 to 50 years old were excluded).

Predictive Ability

The model demonstrated a sensitivity of 75.3% and a specificity of 59.3%. The positive predictive value

Table 3. Variables Associated With Screening Mammography in Women With Intellectual Disabilities–Bivariate Analysis

Variables	Patients (N = 2907)	Patients With Mammogram* (%)	Unadjusted Odds Ratio (95% CI)
Intrapersonal			
Age (years)			
40–49	1022	51	0.89 (0.75–1.05)
50–59	1119	54	Reference
60–69	617	55	1.04 (0.85–1.27)
70–74	149	46	0.73 (0.52–1.03)
Psychiatric diagnosis			
≥1	1785	55	1.44 (1.24–1.67)
≥2	786	57	1.30 (1.10–1.53)
≥3	258	61	1.45 (1.12–1.89)
None			Reference
ADLs (summary score)			
0	816	57	
1	485	54	Reference
2	377	53	
3	408	46	0.78 (0.66–0.91)
4	585	48	
Requires sedation for clinical visits			
Yes	652	50	0.81 (0.68–0.96)
No	2070	56	Reference
Requires special positioning for exams			
Yes	159	48	0.73 (0.53–1.00)
No	2496	56	Reference
Uncooperative or requires limited waiting period			
Yes	725	50	0.77 (0.65–0.92)
No	1921	57	Reference
Down syndrome			
Yes	383	43	0.64 (0.53–0.80)
No	2524	54	Reference
Family history of breast cancer			
Yes	212	67	1.85 (1.37–2.48)
No	2695	52	Reference
Interpersonal			
Guardian assigned			
Yes	1811	50	0.76 (0.65–0.89)
No	1096	57	Reference
Able to communicate			
Yes	1780	57	1.54 (1.32–1.78)
No	1113	47	Reference
System level			
Residential setting			
24-hour support	1700	56	1.32 (1.14–1.53)
Not 24-hour support	1207	49	Reference
Health coordination by RN			
Yes	1525	57	1.40 (1.21–1.63)
No	1382	48	Reference
Colon cancer screening (age ≥50 years)			
Yes	761	68	2.18 (1.73–2.73)
No	539	49	Reference

(*Continued*)

Table 3. Continued

Variables	Patients (N = 2907)	Patients With Mammogram* (%)	Unadjusted Odds Ratio (95% CI)
Bone density screening (age ≥50 years)			
Yes	943	68	2.46 (1.90–3.17)
No	334	46	Reference
Ever had Pap or GYN exam			
Yes	1935	64	3.81 (3.24–4.49)
No	972	31	Reference
Flu vaccine given 2007 or after			
Yes	1690	68	4.38 (3.74–5.12)
No	1217	32	Reference

*Mammogram occurred between January 1, 2007, and December, 31 2008.
ADL, activity of daily living; RN, registered nurse; GYN, gynecologic; Pap, Papanicolaou smear.

was 70.2% and the negative predictive value was 65.4%.

Discussion

There are few data about screening mammography in the United States among women with intellectual disabilities. These data indicate an overall rate of screening within the past 2 years of 53%. This is higher than other non-US populations of women with intellectual disabilities but much lower than the rate of 84.9% found in the general population in Massachusetts.[12] These data show several individual and system-level variables positively associated with mammography in intellectually disabled women: living in 24-hour supported residential settings, having health coordination by a nurse, having a family history of breast cancer, receiving the influenza vaccine (a likely marker for preventive care), and communication ability. Though not all these variables are modifiable, several have been associated with preventive care in other studies. Some variables were negatively associated with mammography: having a guardian, Down syndrome, or higher levels of ADL needs. In the sensitivity analysis examining only subjects living in 24-hour residential settings, ADL needs and having a guardian disappeared from the final model.

The association of health coordination by a nurse with mammography (and particularly the interaction between health coordination and special needs relative to the examination) underscores the potential of an RN already involved with the subject to positively advocate for them to receive preventive services. Though few rigorous studies have analyzed the impact of health coordination on health care for people with intellectual disabilities,[24] the relationship has been noted indirectly. For example, researchers note that nurses play an important role in facilitating access to breast cancer screening for women with intellectual disabilities[25–27] in terms of both helping their clients overcome barriers to screening and the effect of their own knowledge about screening on their clients' screening patterns. During health coordination activities, it is likely that the RN prompts the health care provider to consider a mammogram and then problem-solves the logistic aspects of getting the test for the client (ie, calling the mammography center to reserve extra time or ensuring that women who require sedation are adequately medicated and the staff are prepared for the experience). For older women in the general population, researchers have noted that practice-level factors[28] and relationship-centered aspects of the medical home[29] affect preventive screening, again pointing to the potential for a health care professional to advocate for preventive services.

Though the population of women living outside settings with 24-hour support was not as well represented in this study, the above findings likely have significance for this group as well. We suspect that women with intellectual disabilities who live more independently in the community or with family are less able to consistently access preventive care. They may also receive advice and assistance from family members who are not as well-informed about prevention as the RN providing health care coordination would be. For example, having a guardian was associated with a lower likelihood of

Table 4. Logistic Regression Showing Adjusted Association With Mammography*

Variable	Adjusted Odds Ratio (95% CI)†
Guardian assigned	
Yes	0.77 (0.61–0.95)
No	Reference
Down syndrome	
Yes	0.63 (0.48–0.82)
No	Reference
Able to communicate	
Yes	1.44 (1.14–1.81)
No	Reference
Requires special positioning for examinations	
Yes	0.65 (0.44–0.95)
No	Reference
Family history of breast cancer	
Yes	1.91 (1.35–2.70)
No	Reference
Flu vaccine given 2007 or after	
Yes	4.67 (3.84–5.66)
No	Reference
Interaction between uncooperative/ requires limited waiting period at medical exams and health coordination by RN	
Health coordination by RN	
Uncooperative or requires limited waiting period	0.92 (0.81–1.05)
Cooperative and does not require limited waiting period	Reference
No health coordination by RN	
Uncooperative or requires limited waiting period	0.79 (0.71–0.89)
Cooperative and does not require limited waiting period	Reference
Interaction between residential setting and ADL	
Receives 24-hour residential support	
High ADL score	0.88 (0.78–1.01)
Low ADL score	Reference
Receives less than 24-hour residential support or no support	
High ADL score	0.77 (0.68–0.87)
Low ADL score	Reference

*C statistic = 0.728.
†Final model.
ADL, activity of daily living; RN, registered nurse.

mammography, except in the population of women living in settings with 24-hour support.

For women living outside these residential settings, the issue of how to approximate health care coordination and improve access is not easily resolved. One potential solution would be to shift that responsibility to the health care provider, requesting that all primary care practices review the prevention and screening practices for vulnerable patients (potentially extending beyond women with intellectual disabilities), facilitating their involvement in screening and prevention. The patient-centered medical home movement may be an excellent initiative to develop practice-based procedures and/or pilot interventions around this issue. However, these potential solutions do not address the issue of women with intellectual disabilities in the community who do not receive consistent primary care.

An interesting and somewhat counterintuitive finding was the association of higher numbers of psychiatric diagnoses with mammography. Although this finding is preliminary (based on secondary data analyses), one potential explanation is that women with psychiatric diagnoses in their record probably receive care and medication for these

diagnoses, potentially affecting their ability to tolerate the anxiety of mammography.

In addition, subjects with high ADL support needs (requiring assistance in at least three domains out of four) who did not receive 24-hour residential supports were less likely to receive mammography. It is unknown whether this is reflective of a more medically complex, fragile group who may not represent good candidates for screening and preventive care versus a group overwhelmed by the logistic difficulties of getting some of these patients to the examination. However, because this barrier seems to be ameliorated by the involvement of 24-hour residential supports, it is likely at least some of these subjects represent people who are good candidates for screening but experience logistic challenges. Researchers have noted health disparities among people with disabilities who have relatively more functional impairments[30] and an increased likelihood of preventive care for people with intellectual disabilities who live in 24-hour residential settings.[31] Future research should determine whether the high ADL support needs generally represent a person who may not be considered eligible for screening versus someone who is eligible but is not receiving mammograms.

It was also intriguing to note low rates of mammography among women with Down syndrome. There are scant US data on this topic, but European researchers have suggested that the breast cancer risk is so low for women with Down syndrome that they are actually at higher risk of radiation injury from mammography[20] and should be counseled not to have routine mammography. It is unclear whether the low rates among women with Down syndrome in our population reflect application of this recommendation by US physicians. It has not been shown, however, that there is a significant risk of radiation injury from mammography for women without Down syndrome.[32]

This study had several limitations. Because mammography is not a rare event, the ORs presented here are higher than a comparable rate ratio would be; ORs were used to be consistent with other, similar studies. The database, though highly representative of women with intellectual disabilities who live in supported settings, has lower representation of women who live with families or in the community without state supports. Therefore, generalizing to the entire population of intellectually disabled women is not possible. Second, this database was designed for other purposes and lacked certain variables that are usually considered, ie, race, ethnicity, and level of education. Third, some records may have underreporting of certain disabilities or medical conditions. However, these misclassifications are not suspected to be biased with regard to mammography screening. Fourth, the database lacked information about obesity, which is known to be common in people with intellectual disabilities[33–35] and to be associated with lower rates of screening for some cancers.[36] Fifth, because the study was conducted in Massachusetts, a state that has universal health insurance, we were unable to assess the impact of lack of insurance coverage on the likelihood of mammography. Despite these limitations, this database is large, only includes intellectually disabled women, and yielded results that confirmed the model's validity.

Several Federal initiatives[5,6] have encouraged providers and health systems to improve primary and preventive care for adults with intellectual disabilities. These data indicate potential areas for intervention; at the system level, health coordination could be broadened or made available to more clients, and guardians could be targeted for more education about screening and health recommendations for people with intellectual disabilities. At the provider level, women with intellectual disabilities who do not live in supported settings could be particularly vulnerable and should be educated and supported in pursuing breast cancer screening.[37] Primary care physicians also should be aware of the extent to which residential setting can determine prevention and screening opportunities for people with intellectual disabilities. These findings should be helpful in increasing awareness of characteristics associated with lower rates of screening and prevention for members of a vulnerable, underserved population present in many community primary care practices.

The authors thank Pamela Ohman Strickland, MS, PhD, at the University of Medicine and Dentistry of New Jersey School of Public Health, for her review and comments on the statistical methods of this article. We also thank the Massachusetts Department of Developmental Services for providing access to data.

References

1. Chew KL, Iacono T, Tracy J. Overcoming communication barriers – working with patients with intellectual disabilities. Austr Fam Phys 2009;38(1–2):10–4.

2. Anderson D, Lakin K, Bruininks R, Hill B. A national study of residential and support services for elderly persons with mental retardation. Minneapolis, MN: Center for Residential and Community Services, University of Minnesota; 1987.
3. Iacono T, Sutherland G. Health screening and developmental disabilities. J Pol Prac Intell Disabil 2006;3(3):155–63.
4. Wilkinson JE, Cerreto MC, Culpepper LC. Primary care screening tests for adults with intellectual disabilities. J Am Board Fam Med 2007;20(7):399–407.
5. US Department of Health and Human Services. Healthy People 2010. Available at: www.healthypeople.gov [homepage]. Accessed 25 September 2011.
6. US Department of Health and Human Services. Closing the Gap: A National Blueprint to Improve the Health of Persons with Mental Retardation. Report of the Surgeon General's Conference on Health Disparities and Mental Retardation. Available at: http://www.surgeongeneral.gov/topics/mentalretardation/retardation.pdf. Accessed 27 February 2007.
7. Wilkinson JE, Cerreto MC. Primary care for women with intellectual disabilities. J Am Board Fam Med 2008;21:215–22.
8. Sullivan SG, Hussain R, Threlfall T, Bittles AH. The incidence of cancer in people with intellectual disabilities. Cancer Cause Control 2004;15:1021–5.
9. Patja K, Eero P, Iivanainen M. Cancer incidence among people with intellectual disability. J Intell Disabil Res 2001;45(Pt 4):300–7.
10. Center for Developmental Disabilities Evaluation and Research, Commonwealth of Massachusetts Executive Office of Health and Human Services, Department of Mental Retardation. 2007 mortality report. Available at: http://www.mass.gov/Eohhs2/docs/dmr/mortalityreport2007.pdf. Accessed 21 September 2010.
11. Cancer Incidence and Mortality in Massachusetts, http://www.mass.gov/Eeohhs2/docs/dph/cancer/registry_statewide_03_07_report.pdf. Accessed 7 October 2011.
12. Centers for Disease Control and Prevention. Behavioral risk factor surveillance system. Prevalence and trends data. Massachusetts–2008 women's health: women aged 40+ who have had a mammogram within the past two years. Available at http://apps.nccd.cdc.gov/brfss/display.asp?cat=WH&yr=2008&qkey=4421&state=MA. Accessed 11 October 2010).
13. Davies N, Duff M. Breast cancer screening for older women with intellectual disability living in group homes. J Intellect Disabil Res 2001;45(Pt 3):253–7.
14. Sullivan SG, Glasson EJ, Hussain R, et al. Breast cancer and the uptake of mammography screening services by women with intellectual disabilities. Prev Med 2003;37(5):507–12.
15. Wisdom JP, McGee MG, Horner-Johnson W, Michael YL, Adams E, Berlin M. Health disparities between women with and without disabilities: a review of the research. Soc Work Pub Health 2010;25(3):368–86.
16. Luckasson R, Coulter DL, Polloway EA, et al. Mental Retardation: Definition, Classification and Systems of Support. Washington, DC: American Association of Mental Retardation; 2002.
17. McLeroy K, Bibeau D, Stecker A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(35):351–77.
18. Wilkinson JE, Bowen DJ, Freund K, Rosen AK. Evaluating cancer screening outcomes in a Department of Mental Retardation Database. Presented at: North American Primary Care Research Group, Montreal, Canada, November 2009.
19. US Preventive Services Task Force. Screening for breast cancer. 2009. Available at: http://www.uspreventiveservicestaskforce.org/uspstf/uspsbrca.htm. Accessed 16 September 2010.
20. Yang O, Rasmussen SA, Friedman JM. Mortality assoc with Down's syndrome in the USA from 1983 to 1997: a population-based study. Lancet 2002; 359(9311):1019–25.
21. Satge D. Breast cancer screening guidelines should be adapted in Down syndrome. Br Med J 2002; 324(7346):1155.
22. MMWR Quick Guide. Recommended adult immunization schedule—United States, 2010. Available at: http://www.cdc.gov/mmwr/PDF/wk/mm5901-Immunization.pdf. Accessed 16 September 2010.
23. Shtatland S, Kleinman K, Cain EM. A new strategy of model building in Proc Logistic with automatic variable selection, validation, shrinkage and model averaging. SAS Users Group International (SUGI) 29 Proceedings. Montréal, Canada, May 9–12, 2004.
24. Balogh R, Oullette-Kuntz H, Bourne L, Lunsky Y, Colantonio A. Organising health care services for persons with an intellectual disability. Cochrane Database Systems Rev 2008;(4):CD007492.
25. McIlfatrick S, Taggart L, Truesdale-Kennedy M. Supporting women with intellectual disabilities to access breast cancer screening: a healthcare professional perspective. Eur J Cancer Care (Engl) 2011;20:412–20.
26. Kirby S, Hegarty J. Breast awareness within an intellectual disability setting. Eur J Oncol Nurs 2010; 14(4):328–36.
27. Tyler CV, Zyzanski SJ, Panaite V, Council L. Nursing perspectives on cancer screening in adults with intellectual and other developmental disabilities. Intellect Dev Disabil 2010;48(4):271–7.
28. Pham HH, Schrag D, Hargraves JL, Bach PB. Delivery of preventive services to older adults by primary care physicians. JAMA 2005;294(4):473–81.
29. Ferrante JM, Balasubramanian BA, Hudson SV, Crabtree BF. Principles of the patient-centered medical home and preventive services delivery. Ann Fam Med 2010;8(2):108–16.
30. Henderson CM, Rosasco M, Robinson LM, et al. Functional impairment severity is associated with

health status among older persons with intellectual disability and cerebral palsy. J Intellect Disabil Res 2009;53(11):887–97.
31. Bershadsky J, Kane RL. Place of residence affects routine dental care in the intellectually and developmentally disabled adult population on Medicaid. Health Serv Res 2010;45(5 Pt 1):1376–89.
32. Feig SA, Hendrick RE. Radiation risk form screening mammography of women aged 40–49 years. J Natl Cancer Inst Monogr 1997;(22):119–124.
33. Moran R, Drane W, McDermott S, Dasari S, Scurry JB, Platt T. Obesity among people with and without mental retardation across adulthood. Obes Res 2005; 13(2):342–9.
34. Yamaki K. Body weight among adults with intellectual disability in the community. Ment Retard 2005; 43(1):1–10.
35. Rimmer JH, Braddock D, Fujiura G. Prevalence of obesity in adults with mental retardation: implications for health promotion and disease prevention. Ment Retard 1993;31(2):105–10.
36. Ferrante JM, Chen PH, Jacobs A. Breast and cervical cancer screening in obese minority women. J Womens Health (Larchmt) 2006;15(5):531–41.
37. Wilkinson JE, Deis C, Bowen DJ, Bokhour BG. "It's easier said than done": perspectives on mammography from women with intellectual disabilities. Ann Fam Med 2011;9(2):142–7.

Bone 48 (2011) 307–311

Contents lists available at ScienceDirect

Bone

journal homepage: www.elsevier.com/locate/bone

Height loss, vertebral fractures, and the misclassification of osteoporosis

WanWan Xu [a], Subashan Perera [a], Donna Medich [a], Gail Fiorito [a], Julie Wagner [a], Loretta K. Berger [b], Susan L. Greenspan [a,*]

[a] Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA
[b] Department of Family Medicine, University of South Alabama College of Medicine, Mobile, AL, USA

ARTICLE INFO

Article history:
Received 12 July 2010
Revised 1 September 2010
Accepted 15 September 2010
Available online 24 September 2010

Edited by: Robert Recker

Keywords:
Vertebral fractures
Vertebral Fracture Assessment
Osteoporosis
Bone mineral density

ABSTRACT

Background: The presence of a vertebral fracture identifies a patient who has clinical osteoporosis. However, approximately 2/3 to 3/4 of VFs are asymptomatic. Vertebral Fracture Assessment is a method derived from dual-energy X-ray absorptiometry (DXA) to assess vertebral fractures. The objectives of this study were 1) to determine the association between the degree of height loss in older men and women and the risk of a vertebral fracture, and 2) to determine if the knowledge of vertebral fractures will alter the classification of osteoporosis based on bone mineral density alone.
Methods: 231 men and women over the age of 65 underwent DXA scan of their spine and hip (including bone mineral density and Vertebral Fracture Assessment), measurement of their height, and a questionnaire.
Results: We found that height loss was significantly associated with a vertebral fracture ($p = 0.0160$). The magnitude of the association translates to a 19% increase in odds for 1/2 in. and 177% for 3 in. Although 45% had osteoporosis by either bone mineral density or fracture criteria, 30% would have been misclassified if bone mineral density criteria were used alone.
Conclusions: Height loss is an indicator for the presence of vertebral fractures. Bone mineral density criteria alone may misclassify older patients who have osteoporosis.

Introduction

In addition to low bone mass, a previous fracture significantly increases the risk of developing a future fracture. A vertebral fracture increases the risk of a new fracture up to 4-fold [1–4]. In addition to needing therapy, patients with vertebral fractures may suffer from increased back pain, increased days of bed rest, decreased time at work, depression, and a higher rate of hospitalization and mortality [5,6]. Thus, the presence of a vertebral fracture has a significant clinical impact.

Approximately two-thirds to three-quarters of vertebral fractures are asymptomatic [7]. Diagnosis of an asymptomatic vertebral fracture is not common in clinical practice, because it requires a lateral X-ray, often performed at a different facility (hospital or radiology unit) and increases the patient's expense and radiation exposure. By contrast, Vertebral Fracture Assessment (VFA) is a dual-energy X-ray absorptiometry (DXA)-based technology that can be performed at the same point of service during bone density assessment and uses a much lower radiation dose than a standard lateral X-ray. Greenspan et al. has previously found that vertebral fractures assessed by VFA were present in 18.3% of asymptomatic, postmenopausal women who were screened for osteoporosis by DXA [8]. When using BMD to classify osteoporosis using the World Health Organization classification, sensitivity ranged from 40 to 74%. When VFA was included to classify patients with osteoporosis, as much as 60% of osteoporotic individuals would have been missed by BMD alone. Therefore, VFA may be a useful adjunct in the clinical identification of osteoporosis and may prevent mismanagement of osteoporotic patients.

The International Society of Clinical Densitometry has suggested indications for VFA including women who have had a height loss from the peak height of ≥ 1.6 in., men who have had a height loss of ≥ 2.4 in., or patients with a prospective height loss of >0.8 in. in women or >1.2 in. in men [9]. However additional data are needed to support these recommendations. The objective of this study was to determine the association between the degree of height loss in both older men and women and the risk of a vertebral fracture, and to determine if VFA will increase the prevalence of osteoporosis diagnosis in men and women who are being screened for osteoporosis by BMD alone.

Material and methods

Study subjects

Two hundred and thirty-one men and women over the age of 65 were recruited from the Claude D. Pepper Registry, Benedum Geriatric Clinic at the University of Pittsburgh Medical Center, Braddock

* Corresponding author. Kaufmann, 1110, Osteoporosis Prevention and Treatment Center, 1110 Kaufmann Building, 3471 Fifth Avenue, Pittsburgh, PA 15213, USA. Fax: +1 412 692 2486.
E-mail address: greenspans@dom.pitt.edu (S.L. Greenspan).

8756-3282/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.bone.2010.09.027

Hospital, and through advertising. The Claude D. Pepper Registry is a database of more than 2000 Pittsburgh participants over the age of 60 who are willing to participate in research. Reasons for exclusion were morphological abnormalities, metabolic bone disease other than osteoporosis, and consumption of medication that affect bone mineral metabolism positively or negatively other than calcium and vitamin D. The study was approved by the institutional review board at the University of Pittsburgh Medical Center, PA, and all participants signed written informed consent.

Outcome variables

Bone mineral densitometry of the PA spine and hip were assessed by dual-energy X-ray absorptiometry (DXA) using a Hologic Discovery A bone densitometer located in the Benedum Geriatric Clinic and the Clinical Translational Research Center at the University of Pittsburgh Medical Center. Measurements were obtained and analyzed using standard manufacturer protocols. World Health Organization (WHO) criteria were used to classify bone health and osteoporosis [10]. A T-score of −1.0 SD and above is considered normal, a T-score between −1.0 SD and −2.5 SD is considered low bone mass (osteopenia), and a T-score of −2.5 SD and below is considered osteoporotic [11]. Bone mineral density measurements were taken for the spine, total hip, and the femoral neck as suggested by recommendations of the International Society for Clinical Densitometry [12]. The site with the lowest T-score was included in the classification of osteoporosis.

VFA was performed through lateral spine imaging of T4–L4 on Hologic Discovery A using manufacturer standard protocols. The vertebral bodies from the scan were visually inspected by a technician and fractures were classified using the method of Genent (semiquantitative visual assessment) with the assessment of computer-calculated reduction in vertebral height [13]. Two technicians trained in VFA examined VFA results independently and were blinded to the BMD results. If differences were noted, the VFA was reanalyzed by both technicians together.

Clinical characteristics

A questionnaire was administered that contained medical, surgical, gynecological, fracture, and family history in addition to medications, alcohol, tobacco, and exercise. Participants were also asked to report their tallest height ever from memory. Nontraumatic fragility fractures were defined as fractures that occurred from a fall from standing height or lower. Known traumatic nonvertebral fractures were excluded.

Current height was measured 3 times with a Harpenden stadiometer and an average was calculated. Participants' height loss was calculated as tallest height ever minus current height.

Statistical analysis

The presence of vertebral fractures was operationally defined as a dichotomous variable indicating one or more and three or more fractures. We compared characteristics of subjects with and without VF using independent samples *t*- and chi-square tests, as appropriate, for continuous and categorical measures. We used logistic regression models with the presence of vertebral fractures as the dependent variable, height loss as the primary predictor variable of interest, and exercise, smoking, and calcium/vitamin D intakes covariates. The logistic regression coefficient for height loss was scaled appropriately and then exponentiated to obtain odds ratios for magnitudes of risk increases corresponding to a height loss between 1/2 in. and 3 in. We used a two-way contingency table cross tabulation of subjects meeting criteria of osteoporosis using BMD alone by VF alone. All analyses were performed both with and without stratification by gender. We used SAS® version 9.2 (SAS Institute, Inc., Cary, North Carolina) for all statistical analyses.

Table 1
Clinical characteristics.

	Total	Men	Women
# Participants (%)	231	124 (54%)	107 (46%)
Age (years)	75 ± 6	75 ± 6	75 ± 7
Average height loss (inches)	1.86 ± 1.22	1.93 ± 1.26	1.77 ± 1.17
Average % height loss	2.75% ± 1.79%	2.75% ± 1.79%	2.75% ± 1.80%
Average BMI (kg/m^2)	29 ± 5	28 ± 5	29 ± 6
# with 1+ vertebral fracture (%)	91 (39%)	45 (36%)	46 (43%)
# with 3+ vertebral fracture (%)	20 (9%)	9 (7%)	11 (10%)
History of fragility fractures	49 (21%)	15 (12%)	34 (32%)
Smoking (%)	116 (50%)	82 (66%)	34 (32%)
Exercise (%)	177 (77%)	111 (90%)	66 (62%)
Calcium intake	134 (58%)	64 (52%)	70 (65%)
Vitamin D intake	134 (58%)	64 (52%)	70 (65%)
Results: mean ± SD or N (%)			

Results

We recruited 231 participants, including 124 men and 107 women. The average age was 75 ± 6 years and their average height loss from peak was 1.86 ± 1.22 in. (mean ± SD). Ninety-one participants had one or more vertebral fractures and 20 had three or more vertebral fractures. Twenty-one percent had a previous history of an adult fragility fracture (including 3 with a hip fracture), 77% participated in exercise, and 58% took calcium and vitamin D supplements. There were no significant differences between men and women in terms of their vertebral fracture rate (Table 1). However, those who exercised were less likely to have one or more vertebral fractures than those who did not (35% with exercise vs 54% with no exercise, $p = 0.014$). Furthermore, those with 3 or more vertebral fractures were less likely to take calcium and vitamin D (5%) than those who did not (13%, $p = 0.029$) (Table 2).

Participants with one or more vertebral fractures had greater height loss (mean 2.18 ± 1.30 in.) than subjects without vertebral fractures (1.64 ± 1.11 in.). Forty-three percent of women had one or more VF and 36% of men had one or more vertebral fractures. In women, those with no VF had an average height loss of 1.54 in., whereas those with one or more vertebral fractures had lost an average of 2.07 in. ($p = 0.0184$). In men, those with no vertebral fracture had an average height loss of 1.72 in., whereas those with one or more vertebral fractures had lost an average of 2.30 in.

Table 2
Rates of having VFs based on exercise, smoking, and calcium/vitamin D intake.

		Rate of having 1 or more VFs	P value	Rate of having 3 or more VFs	P value
All	Exercise	35%	0.014	7%	>0.05
	No exercise	54%		13%	
	Smoking	38%	>0.05	7%	>0.05
	No smoking	41%		10%	
	Ca/VitD	34%	>0.05	5%	0.029
	No Ca/VitD	46%		13%	
Women	Exercise	30%	0.0008	6%	>0.05
	No exercise	63%		17%	
	Smoking	38%	>0.05	6%	>0.05
	No smoking	45%		12%	
	Ca/VitD	40%	>0.05	9%	>0.05
	No Ca/VitD	49%		14%	
Men	Exercise	38%	>0.05	8%	>0.05
	No exercise	23%		0%	
	smoking	38%	>0.05	7%	>0.05
	No smoking	33%		7%	
	Ca/VitD	28%	0.0508	2%	0.0116
	No Ca/VitD	45%		13%	

Table 3
Height loss and odds of having a vertebral fracture.*

Unit of height loss (inches)	Odds ratio	95% confidence limits
0.5	1.19	1.06–1.34
1	1.40	1.12–1.79
1.5	1.66	0.71–2.15
2	1.97	0.64–2.77
2.5	2.34	0.57–3.58
3.0	2.77	0.51–4.61

P<0.05.
* Adjusted for smoking, exercise, and vitamin D intake.

(p = 0.0146). For each additional 0.5 in. of height loss, the odds ratios for having one or more vertebral fractures increased significantly when adjusted for smoking, exercise, calcium, and vitamin D intake (OR = 1.19, 1.40, 1.66, 1.97, 2.34, 2.77, p<0.05) (Table 3).

According to WHO classification, 60 out of 107 women (56%) from our study had low bone mass and 23 out of 107 (22%) had osteoporosis, compared to 64 out of 124 (52%) and 10 out of 124 (8%) for men, respectively (Table 4). The prevalence of osteoporosis varied when defined by BMD site only vs. BMD site and vertebral fractures combined. For example, 29 out of 231 participants (12.6%) were categorized as osteoporotic when using femoral neck BMD as the sole criteria, but 101 out of 231 participants (43.7%) were categorized as osteoporotic when using femoral neck BMD and vertebral fracture as criteria (Table 4).

In all our participants, 91 out of 231 (39%) had one or more vertebral fractures, and 20 out of 231 (9%) had three or more vertebral fractures (Table 1). Twenty-eight out of 74 subjects (38%) who were categorized as normal by BMD and 42 out of 124 subjects (34%) who were categorized as having low bone mass had one or more vertebral fractures (Fig. 1). In patients with one or more vertebral fractures, the distribution of a normal or low bone mass classification was similar in men and women. One hundred and three out of 231 patients (45%) had osteoporosis by BMD classification or the presence of a vertebral fracture (Table 5). If BMD classification was used alone, 70 out of 231 total participants (30%) would have been misclassified (Table 5). If we exclude 3 patients who had a hip fracture and who should have been classified with clinical osteoporosis by NOF guidelines, in the remaining 228 participants, 69/228 or 30% would have been misclassified.

Table 4
Prevalence of osteoporosis by WHO BMD classification.

		Spine only	Total hip only	Femoral neck only	Any site
All	BMD with vertebral fracture (true classification)				
	Osteoporosis	42.4%	41.1%	43.7%	45.0%
	Low bone mass	16.5%	17.7%	33.8%	35.1%
	Normal	41.1%	41.1%	22.5%	19.9%
	BMD alone				
	Osteoporosis	6.5%	3.5%	12.6%	14.7%
	Low bone mass	26.4%	34.2%	52.4%	53.2%
	Normal	67.1%	62.3%	35.1%	32.0%
Men	BMD with vertebral fracture (true classification)				
	Osteoporosis	38.7%	36.3%	37.1%	38.7%
	Low bone mass	12.9%	13.7%	35.5%	35.5%
	Normal	48.4%	50.0%	27.4%	25.8%
	BMD alone				
	Osteoporosis	4.0%	1.6%	5.6%	8.1%
	Low bone mass	20.2%	25.8%	52.4%	51.6%
	Normal	75.8%	72.6%	41.9%	40.3%
Women	BMD with vertebral fracture (true classification)				
	Osteoporosis	46.7%	46.7%	51.4%	51.4%
	Low bone mass	20.6%	22.4%	31.8%	35.5%
	Normal	32.7%	30.8%	16.8%	13.1%
	BMD alone				
	Osteoporosis	9.3%	5.6%	20.6%	21.5%
	Low bone mass	33.6%	43.9%	52.3%	56.1%
	Normal	57.0%	50.5%	27.1%	22.4%

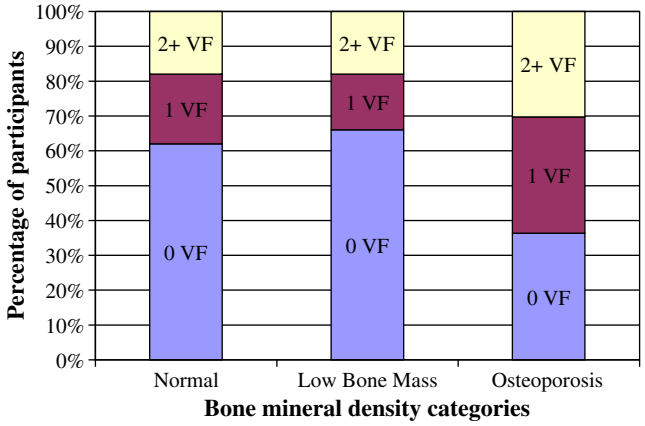


Fig. 1. The percentage of participants classified with normal, low bone mass or osteoporosis by bone mineral density who had zero (0 VF), one (1 VF) or two or more vertebral fractures (2+ VF) by Vertebral Fracture Assessment.

Discussion

We found that height loss was significantly associated with a vertebral fracture. The magnitude of the association translates to a 19% increase in odds of having a vertebral fracture for 0.5 in. of height loss and 177% for 3 in. Those with a vertebral fracture, on average, had a 0.54 in. greater height loss than those without a vertebral fracture. Although 45% had osteoporosis by either BMD or fracture criteria, 30% would have been misclassified if BMD criteria was used alone.

Similar to studies by Siminoski et al., our study showed that the relationship between height loss and prevalent vertebral fractures was significant [14]. Our study also agreed with the cross-sectional studies by Tobias et al., and Gunnes et al. which determined that prevalent vertebral fracture rates increased with greater height loss [15,16]. In the study by Tobias et al. where they included 509 women, the difference in height loss between the two groups was 0.31 in. Our participants with one or more vertebral fractures lost an average height of 0.54 in. more than our participants with no vertebral fractures.

Prospective studies have also demonstrated a correlation between height loss and incident fractures. Moayyeri et al. [17] recruited 25,623 men and women 40–79 years old, measured their height and followed their fracture status over several years. The study showed that height loss is an independent risk factor for osteoporotic fractures. Kaptoge et al. [18] showed that the risk of incident vertebral fracture significantly increased with increased height loss. Some studies have suggested a height loss cutoff to indicate increased risk of vertebral fractures. Due to the small number of participants, we were unable to find a statistically significant threshold value. However, our data did suggest a sharp increase in the number of vertebral fractures once participants have lost more than 3 in. This finding is similar to that of Vallarta-Ast et al. [19] who found an increase in fracture prevalence in men when the historical height loss is >2.5 in. and also to that of Siminoski, who suggested the cutoff to be 6 cm (2.4 in.).

Table 5
Misclassification of osteoporosis.

	No VF	Yes VF	Total
BMD: normal or low bone mass	128	70	198
BMD: osteoporosis	12	21	33
Total	140	91	231

Siminoski conceded that height loss was not a good screening test for vertebral fracture because it had low sensitivity values and high specificity. Combined with other risk factors, such as age and previous history of fractures, the sensitivity of predicting vertebral fractures may be higher [15,20].

Our study differed from others that looked at vertebral fractures and height loss. Unlike the study by Siminoski et al., our study excluded participants with metabolic bone diseases and those taking bone-altering medications. Unlike the study by Tobias et al. and Gunnes et al., our study included both men and women over the age of 65 with no upper age limit. Compared to previous studies, the prevalent vertebral fracture rate in our study was higher: 36% of men and 43% of women had one or more vertebral fractures. When using BMD WHO classification and the presence of vertebral fractures as criteria for diagnosing osteoporosis, as many as 51% of women and 39% of men had osteoporosis. In Tibias et al., UK women 65–76 years old were recruited; 7.3% of the 509 subjects had one or more vertebral fractures. They identified vertebral fractures with standard X-ray, while we used VFA. However, previous studies have documented that vertebral fractures are often missed on X-rays [21]. Our VFA criteria used a quantitative assessment and identified a vertebral fracture with a loss of height of 20% or more. This quantitative assessment differs from the qualitative technique used by many radiologists. The prevalence of osteoporosis in our population was also higher than those reported by Cheng et al. who reviewed Medicare claims data of 911,327 beneficiaries over the age of 65. Cheng's paper estimated 41.8% of Caucasian females between 75–79 years have osteoporosis, and 8.4% Caucasian males between 75–79 years have osteoporosis. In addition to the quantitative technique, our higher rates of osteoporosis may be explained by the fact that many of our patients were recruited from Benedum Geriatric Clinic. Patients who visit the clinic tend to be more frail, more likely to have fractures, and be osteoporotic. Furthermore, we obtained VFA to assess for vertebral fractures while Medicare claim data could miss asymptomatic fractures. Our high male osteoporotic rate may also be due to the fact that only a minority of vertebral fractures in men comes to clinical attention, contributing to lower rates seen through insurance claims.

Our study demonstrated that men and women with vertebral fractures were often misclassified as having normal or low bone mass by BMD alone. In a study by Freitas et al., 5995 men were recruited and out of those with incident vertebral fractures, 41% had normal BMDs and 46% were categorized as having low bone mass [22]. Similarly, the Rotterdam study by Schuit et al. showed that many women and men with nonvertebral fractures were categorized as normal or having low bone mass by BMD alone [23]. They concluded that BMD needs to be used with other predictors of fractures — such as height loss — to increase its sensitivity. We suggest that in addition to clinical risk factors such as height loss, the VFA is a useful adjunct to BMD and may prevent misdiagnosis of osteoporosis in men and women.

Because this is a cross-sectional study, we cannot determine cause and effect. Although we observed that those who exercised were less likely to have vertebral fractures than those who did not exercise, patients with fractures may be less likely or able to exercise after they fracture. We observed that patients with 3 or more fractures were less likely to take calcium and vitamin D. However due to other comorbid conditions in these patients, they may be less likely to comply with calcium and vitamin D. Furthermore we could not determine if these vertebral fractures were caused by a previous high trauma accident. We did however query participants for known traumatic fractures and only included nontraumatic fragility fractures by their history. A prospective study is needed to corroborate these cross-sectional findings.

This study has many strengths. It is one of the few studies that compare BMD, height loss, and vertebral fractures in both men and women from the same population. Our population included men and women in many stages of health. We recruited healthy, active volunteers as well as frail visitors to the Benedum Geriatric Clinic. Our study also has several limitations. The sample size is small. There may also be selection bias because many of our patients came from Benedum Geriatric Clinic and may be more frail than the general population. Our results may not be applicable to people <65 years old and not applicable to minorities since most of our participants were Caucasian. Moreover our study may be subject to recall bias regarding a patient's previous height. Finally, we did not verify vertebral fractures with a standard X-ray. However, the VFA technique has been validated in the past for moderate and severe vertebral fractures [24,25].

In summary, height loss is an indicator for the presence of vertebral fractures; and the Vertebral Fracture Assessment increases the prevalence of osteoporosis diagnosis in men and women who are being screened for osteoporosis by bone mineral density alone. Those misclassified by BMD alone — 30% of all participants — were correctly classified after having a VFA performed. Further studies are needed to assess longitudinal height loss and incident vertebral fractures assessed by VFA in other clinically relevant cohorts.

Acknowledgments

This study was supported by a gift from the Holleran family, the Osteoporosis Prevention and Treatment Center, Dean's Summer Research Program, T35 DK065521, NIH K24 DK062895, and P30 AG024827 to Dr. Susan Greenspan. Recruitment and data collection were supported by The Pepper Registry, Claude D. Pepper Older Americans Independence Center, Pittsburgh; Benedum Geriatric Center, University of Pittsburgh Medical Center; and the Clinical Translational Research Center, University of Pittsburgh.

References

[1] Melton III LJ, Atkinson EJ, Cooper C, O'Fallon WM, Riggs BL. Vertebral fractures predict subsequent fractures. Osteoporos Int 1999;10:214–21.
[2] Cauley JA, Hochberg MC, Lui LY, Palermo L, Ensrud KE, Hillier TA, et al. Long-term risk of incident vertebral fractures. JAMA 2007;298(23):2761–7.
[3] Klotzbuecher CM, Ross PD, Landsman PB, Abbott III TA, Berger M. Patients with prior fractures have an increased risk of future fractures: a summary of the literature and statistical synthesis. J Bone Miner Res 2000;15:721–39.
[4] van der Klitft M, de Laet CE, McCloskey EV, Johnell O, Kanis JA, Hofman A, et al. Risk factors for incident vertebral fractures in men and women: the Rotterdam Study. J Bone Miner Res 2004;19(7):1172–80.
[5] Nevitt MC, Ettinger B, Black DM, Stone K, Jamal SA, Ensrud K, et al. The association of radiographically detected vertebral fractures with back pain and function: a prospective study. Ann Intern Med 1998;128(10):793–800.
[6] Ensrud KE, Thompson DE, Cauley JA, Nevitt MC, Kado DM, Hochberg MC, et al. Prevalent vertebral deformities predict mortality and hospitalization in older women with low bone mass. J Am Geriatr Soc 2000;48:241–9.
[7] Fink HA, Milavetz DL, Palermo L, Nevitt MC, Cauley JA, Genant HK, et al. What proportion of incident radiographic vertebral deformities is clinically diagnosed and vice versa? J Bone Miner Res 2005 Jul;20(7):1216–22.
[8] Greenspan SL, von Stetten E, Emond SK, Jones L, Parker RA. Instant vertebral assessment: a noninvasive DXA technique to avoid misclassification and clinical mismanagement of osteoporosis. J Clin Densitom 2001;4:373–80.
[9] Baim S, Binkley N, Bilezikian JP, Kendler DL, Hans DB, Lewiecki EM, et al. Official positions of the International Society for Clinical Densitometry and executive summary of the 2007 ISCD Position Development Conference. J Clin Densitom 2008;11(1):75–91.
[10] Kanis JA, for the WHO Study Group. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report. Osteoporos Int 1994;4:368–81.
[11] Kanis JA, Melton III LJ, Christiansen C, Johnston CC, Khaltaev N. The diagnosis of osteoporosis. J Bone Miner Res 1994;9:1137–41.
[12] Anonymous. Official positions of the International Society for Clinical Densitometry. International Society for Clinical Densitometry. West Hartford, CT: International Society for Clinical Densitometry; 2007. Ref Type: Pamphlet.
[13] Genant HK, Wu CY, van Kuijk C, Nevitt MC. Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res 1993;8:1137–48.
[14] Siminoski K, Warshawski RS, Jen H, Lee K. The accuracy of historical height loss for the detection of vertebral fractures in postmenopausal women. Osteoporos Int 2006;17(2):290–6.
[15] Tobias JH, Hutchinson AP, Hunt LP, McCloskey EV, Stone MD, Martin JC, et al. Use of clinical risk factors to identify postmenopausal women with vertebral fractures. Osteoporos Int 2007;18(1):35–43.

[16] Gunnes M, Lehmann EH, Mellstrom D, Johnell O. The relationship between anthropometric measurements and fractures in women. Bone 1996;19(4):407–13.
[17] Moayyeri A, Luben RN, Bingham SA, Welch AA, Wareham NJ, Khaw KT. Measured height loss predicts fractures in middle-aged and older men and women: the EPIC-Norfolk prospective population study. J Bone Miner Res 2008;23(3):425–32.
[18] Kaptoge S, Armbrecht G, Felsenberg D, Lunt M, Weber K, Boonen S, et al. Whom to treat? The contribution of vertebral X-rays to risk-based algorithms for fracture prediction. Results from the European Prospective Osteoporosis Study. Osteoporos Int 2006;17(9):1369–81.
[19] Vallarta-Ast N, Krueger D, Wrase C, Agrawal S, Binkley N. An evaluation of densitometric vertebral fracture assessment in men. Osteoporos Int 2007 Epub ahead of print.
[20] Vogt TM, Ross PD, Palermo L, Musliner T, Genant HK, Black D, et al. Vertebral fracture prevalence among women screened for the Fracture Intervention Trial and a simple clinical tool to screen for undiagnosed vertebral fractures. Fracture Intervention Trial Research Group. Mayo Clin Proc 2000;75(9):888–96.
[21] Ferrar L, Jiang G, Adams J, Eastell R. Identification of vertebral fractures: an update. Osteoporos Int 2005;16(7):717–28.
[22] Freitas SS, Barrett-Connor E, Ensrud KE, Fink HA, Bauer DC, Cawthon PM, et al. Rate and circumstances of clinical vertebral fractures in older men. Osteoporos Int 2008;19(5):615–23.
[23] Schuit SCE, van der Klitft M, Weel AEAM, de Laet CEDH, Burger H, Seeman E, et al. Fracture incidence and association with bone mineral density in elderly men and women: the Rotterdam Study. Bone 2004;34:195–202.
[24] Schousboe JT, DeBold CR. Reliability and accuracy of vertebral fracture assessment with densitometry compared to radiography in clinical practice. Osteoporos Int 2006;17:281–9.
[25] Hospers IC, van der Laan JG, Zeebregts CJ, Nieboer P, Wolffenbuttel BH, Dierckx RA, et al. Vertebral fracture assessment in supine position: comparison by using conventional semiquantitative radiography and visual radiography. Radiology 2009;251(3):822–8.

PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2011; **20**: 192–202
Published online 28 December 2010 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.2088

ORIGINAL REPORT

Impact of implementing alerts about medication black-box warnings in electronic health records†

Donghui Tony Yu[1,2], Diane L. Seger[2], Karen E. Lasser[5], Andrew S. Karson[4], Julie M. Fiskio[2], Andrew C. Seger[1,2,3] and David W. Bates[1,2]*

[1]*Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*
[2]*Department of Information Systems, Partners HealthCare System, Wellesley, MA, USA*
[3]*Massachusetts College of Pharmacy & Health Sciences, Boston, MA, USA*
[4]*Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA*
[5]*Department of Medicine, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA*

ABSTRACT

Background The Food and Drug Administration issues black-box warnings (BBWs) regarding medications with serious risks, yet physician adherence to the warnings is low.

Methods We evaluated the impact of delivering BBW-based alerts about drug–drug, drug–disease, and drug–laboratory interactions for prescription medications in outpatients in an electronic health record with clinical decision support. We compared the frequency of non-adherence to all BBWs about drug–drug, drug–disease, and drug–laboratory interactions for 30 drugs/drug classes, and by individual drugs/drug groups with BBWs between the pre- and post-intervention periods. We used multivariate analysis to identify independent risk factors for non-adherence to BBWs.

Results There was a slightly higher frequency of non-adherence to BBWs after the intervention (4.8% vs. 5.1%, $p = 0.045$). In multivariate analyses, after adjustment for patient and provider characteristics and site of care, medications prescribed during the pre-intervention period were less likely to violate BBWs compared to those prescribed during the post-intervention period (OR 0.67, 95% CI, 0.47–0.96). However, black-box warning violations did decrease after the intervention for BBWs about drug–drug interactions (6.1% vs. 1.8%, $p < 0.0001$) and drug–pregnancy interactions (5.1% vs. 3.6%, $p = 0.01$).

Conclusions Ambulatory care computerized order entry with prescribing alerts about BBWs did not improve clinicians' overall adherence to BBWs, though it did improve adherence for specific clinically important subcategories. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS — drug–drug interactions; drug–laboratory interactions; drug–disease interactions; drug–pregnancy interactions; electronic prescribing; alert; drug safety

Received 26 April 2010; Revised 26 October 2010; Accepted 9 November 2010

INTRODUCTION

Each year, more than 770 000 people are injured or die in hospitals from adverse drug events (ADEs).[1] Reports of in-hospital ADE rates range from 2.4 to 15.0 per 100 admissions[1–5] while in outpatients, the rate of ADEs has been reported to be as high as 27% in a study using direct patient contact to detect ADEs.[6] A 2006 report from the Institute of Medicine on medication errors estimated that at least 1.5 million preventable ADEs occur annually in the United States, with most occurring in the outpatient setting.[7] The high ADE rate found in adult primary care settings is likely due to the high frequency and long duration of exposure to medications in the outpatient population.[6]

Many medication-related injuries are potentially preventable,[2] especially by changing medication-ordering systems.[8] Growing evidence shows that implementing these applications results in benefit; a meta-analysis by Shamliyan *et al.*[9] showed a 66% reduction of prescribing errors on average, and another meta-analysis

* Correspondence to: D. W. Bates, Division of General Internal Medicine and Primary Care, Department of Medicine, Brigham and Women's Hospital, Brigham Circle, 1620 Tremont St., Boston, MA 02120, USA.
E-mail: dbates@partners.org
† Data access and responsibility: Drs Bates and Yu had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

by Wolinsky *et al.*[10] evaluated 10 studies, and found that in 5 ADE rates fell, in 4 there were non-significant trends toward benefit, and in one there was no effect. Most studies evaluating ADE prevention have been conducted in hospital settings; relatively little is known about ADE prevention in the ambulatory setting.[11,12] Through the introduction of advanced computerized prescribing, the medication error rate was reduced by 81% in an adult inpatient population in one study[13] and by 40% in a pediatric inpatient population in another.[14] However, in the outpatient setting, one basic computerized prescribing system did not significantly decrease the frequency of prescribing errors.[15] However, more advanced systems with better decision support should be beneficial in outpatients.[16]

Black-box warnings (BBWs), which are reported among other places in the Physicians' Desk Reference (PDR) are developed by the US Food and Drug Administration (FDA) and are intended to help prescribers avoid the most serious ADEs.[17] Prior studies, including our own,[18] have shown that adherence by prescribers to such warnings is low among ambulatory patients.[19] As a prelude to the present study, we analyzed data on patients from 51 ambulatory practices using one electronic health record (EHR) who received a prescription for a medication in 2002. Of these patients, 10.4% were prescribed a drug with BBWs about drug–drug interactions, drug–disease interactions, drug–lab interactions, or drug–pregnancy interactions, and of patients who were prescribed a BBW drug, 7% received a prescription in violation of a BBW.

In order to improve medication safety and make black-box warnings accessible to clinicians, we added BBW-based drug–drug, drug–disease, drug–lab, and drug–pregnancy alerts to our electronic record. The objective of this study was to evaluate prescribing in violation of BBW before and after delivering BBW-based medication safety alerts to clinicians using ambulatory electronic medical records.

METHODS

Study site

We conducted the intervention at 51 outpatient practices in the greater Boston area that use a common EHR. These included 40 hospital-based clinics, 4 community health centers, and 7 community-based practices. Both primary care and specialty practices were included—many of the hospital based clinics are specialty practices. Most of these practices include non-physician prescribers, the vast majority of prescriptions are written by physicians. We collected information in EHR, known as the longitudinal medical record (LMR) including patient demographics, lists of medical problems and prescription drugs (with start and stop dates), and results of laboratory tests. Partners HealthCare System Institutional Review Board approved this study.

Patient population

We included all patients aged 18 years or older who received at least one prescription medication with a BBW from 1 January to 31 December in 2002 and in 2005. We focused our analysis on patients who received a prescription for a drug that contains a BBW pertaining to drug–drug interactions, drug–laboratory interactions, drug–disease interactions, and/or drug–pregnancy interactions. In the present study, we used the same list of drugs with BBWs as in our previous study.[18] The full list of drugs is available in an online Appendix.

Data collection

For all patients, we collected data on age, gender, race, insurance status, language, medical problems on the problem list (with a date entered on or before 31 December 2005), all medication prescriptions (with start date and/or stop date on or before 31 December 2005), laboratory test results (with report date between 1 January 2001 and 31 December 2002, and between 1 January 2004 and 31 December 2005), type of prescriber (primary care physician, nurse practitioner, physician assistant, specialist physician, and others) and site of care (private offices, hospital-based clinics, or community health centers). Staff at ambulatory practice sites routinely record patient race (white, black, and Asian) and ethnicity (Hispanic); we analyzed these variables to determine if they were associated with prescribing patterns. We defined 'insurance of poverty' as Medicaid or Free Care. We also collected data about alerts and reminders in the LMR at the time of the study. Finally, we reviewed information about BBWs and monitoring recommendations, and any changes in such information between 2002 and 2005, to enable updating of the intervention, but no changes were found.

Interventions

During the intervention, a physician and pharmacist expert panel created a knowledge base. The decision support knowledge database only included alerts

DOI: 10.1002/pds

judged clinically relevant to the ambulatory care setting. We added alerts based on all BBWs regarding drug–pregnancy interactions along with most drug–drug interactions, most drug–lab monitoring, and some drug–disease interactions. We placed each alert into one of four clinical severity tiers.[20,21] Level 1 alerts indicate a fatal or life-threatening interaction. Level 2 alerts indicate an undesirable interaction with the potential for serious injury. Level 3 and level 4 alerts indicate the possibility of an undesirable interaction in which a drug should only be used with caution or may require increased monitoring. All alerts judged sufficiently important were implemented in the LMR.

Prescribing alerts

The LMR generated alerts using information from each patient's active medication list, problem list, and laboratory results, and applied logic rules to identify potential contraindications. The alerts appeared as an on-screen warning that identifies the alerted drug and the issues (examples indicated in Figures 1–3). A single medication could generate multiple alerts displayed on a single screen, each requiring a separate clinician action. These alerts appeared to clinicians as soon as they entered a new medication. Clinicians were interrupted with pop-up boxes displaying the alerts for level 1 and level 2 alerts only, and these required an action to acknowledge the alert (Figures 1 and 2). Level 3 alerts were non-interruptive and no action was required (Figure 3). Other warnings besides the BBW were also displayed. There was no specific notation when a BBW alert was displayed as compared to one of these other non-BBW-based warnings. All alerts were presented in hierarchical order based on their alert level with level 1 being the most severe where the user has to take action and cannot continue with the order followed by level 2 alerts where the user may override the alert but must give a reason followed by level 3 and level 4 alerts where monitoring or follow-up is recommended but is only presented as informational.

Follow-up for lab monitoring alerts

Follow-up lab alerts (Level 4) were non-interruptive and presented in the LMR as reminders. If a patient had been on a medication for more than 12 months and had not had the appropriate labs checked during the 12-month time period, the user was presented with a reminder to check labs. For example, a provider could receive the following alert: 'Pt on Metformin >365 consecutive days. Checking Creatinine (serum) is recommended.'

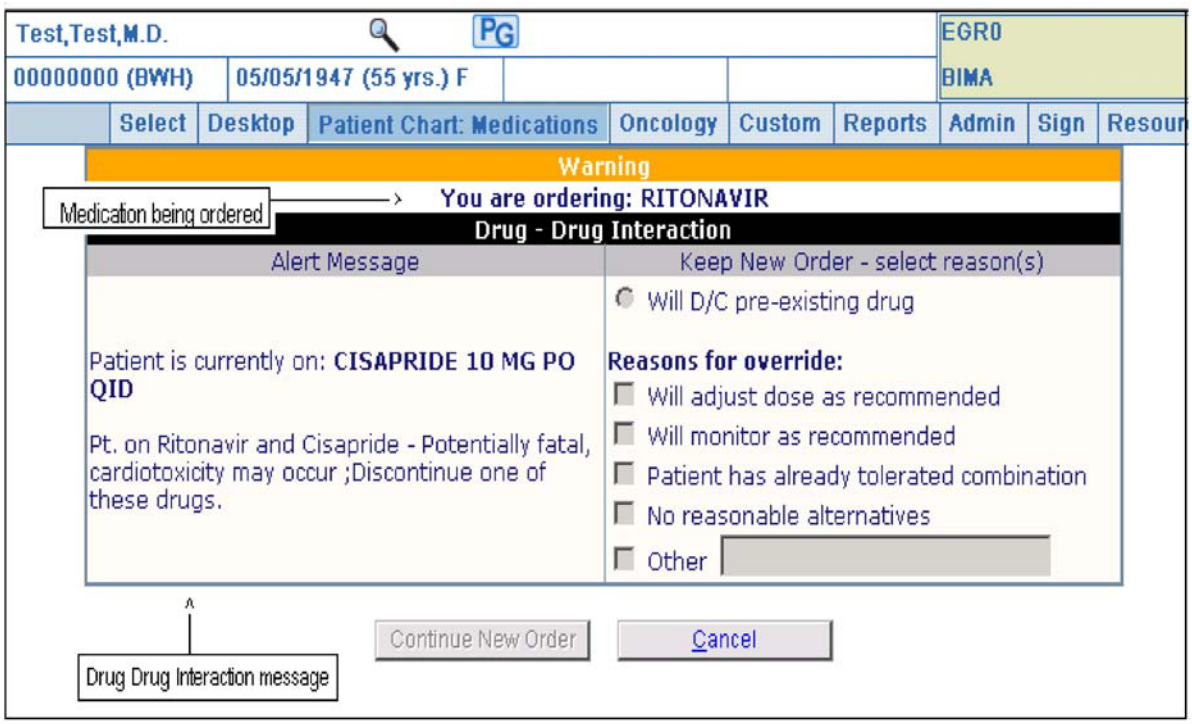


Figure 1. Level 2 alert for drug–drug interaction, the clinician may choose either to keep both orders or stop one of them

DOI: 10.1002/pds

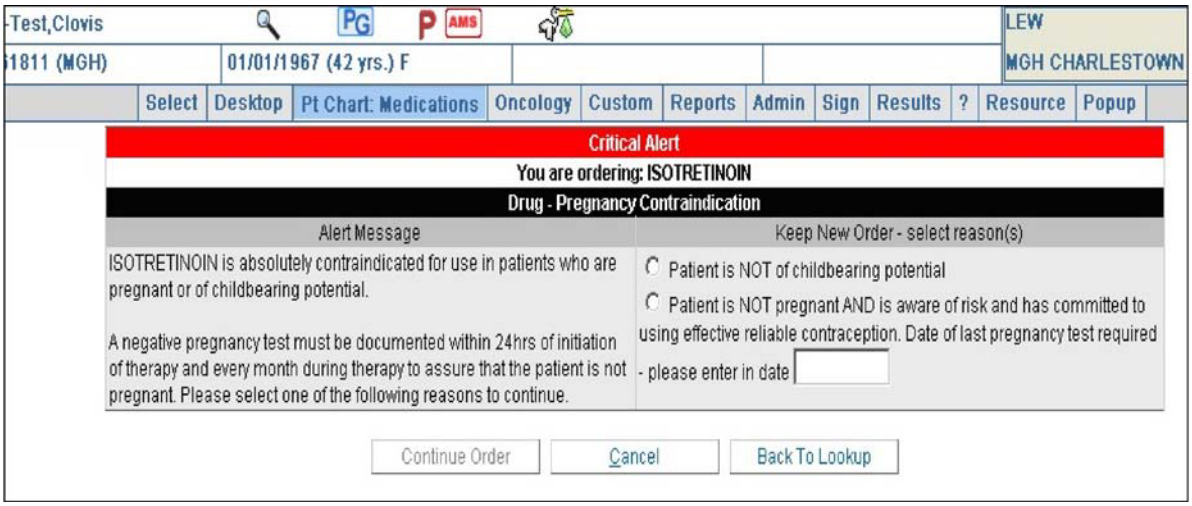


Figure 2. Level 1 alert for drug–pregnancy contraindication, requiring order to be stopped if patient has a positive pregnancy test within the past 9 months or is of childbearing potential

Outcome measures

The primary outcomes were the frequency of non-adherence to BBWs in total; by individual categories of drug–drug, drug–disease, drug–laboratory, and drug–pregnancy warnings; and by each individual drug with a BBW. Secondary outcomes were correlates of non-adherence to BBWs. To measure non-adherence to BBWs, we used the same definitions as those used in our previous study.[18] For example, for a patient starting a drug for which baseline hepatic or renal function was required, we scored the prescribing as adherent if tests for hepatic or renal dysfunction occurred within 3 months before the start of the drug. Similarly, if baseline pregnancy testing was indicated, then clinicians were deemed adherent if a pregnancy test

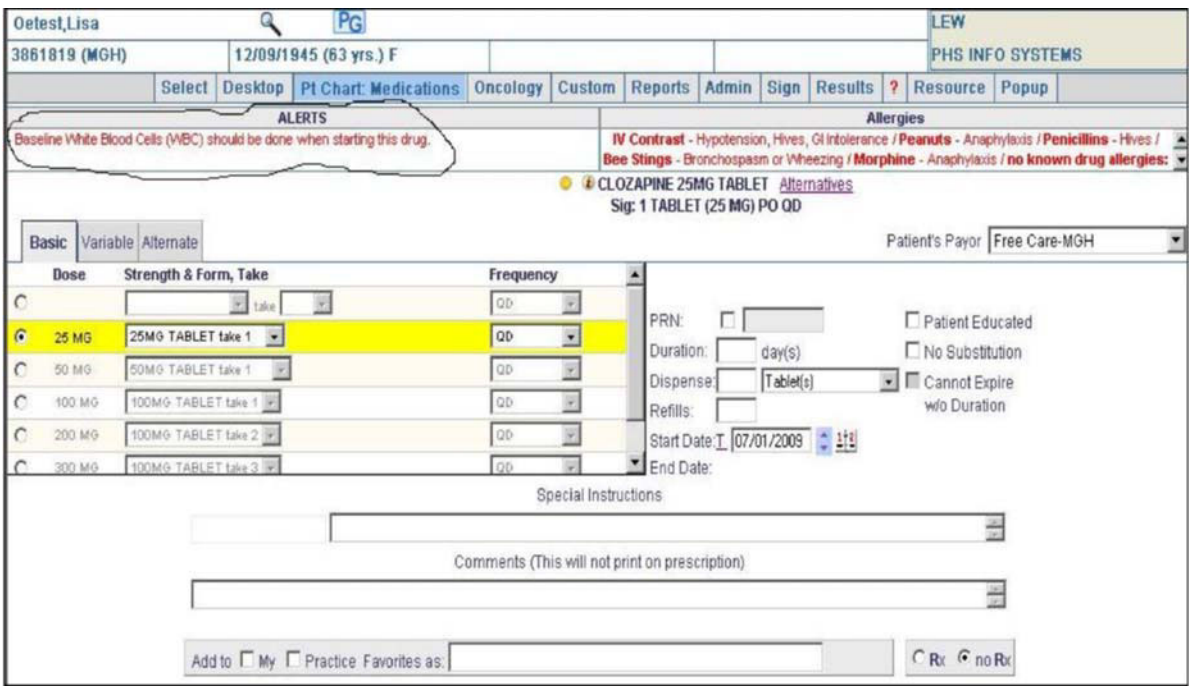


Figure 3. Level 3 alert for drug–laboratory interaction, clinician is reminded that baseline white blood cells (WBC) count should be obtained when starting the drug

DOI: 10.1002/pds

occurred within 1 month before starting the drug. For all other required baseline laboratory tests, we considered the prescribing to be adherent if the tests were performed within 12 months prior to starting the drug. We considered non-adherence to BBWs for drug–drug interactions or drug–disease interactions to be present if either a drug with BBW or the interacting drug was prescribed while the other was administered on the same day, or a medication with a BBW was prescribed to a patient with a disease contraindicating the drug. However, we scored prescribing as adherent if a prescriber entered a discontinuation date for the drug within 3 months of the appearance of a contraindicated drug, disease, or laboratory value. The term 'non-adherence' used throughout the paper has the same meaning as the term 'violation.' Any reason for overriding by a prescriber as the response to an alert was considered to be a violation of a BBW.

Analyses

We conducted before–after comparisons between the pre- (2002) and post-intervention (2005) periods among all medication prescriptions with a BBW and with an alert implemented in the LMR during the intervention period. The unit of analysis was the individual medication prescription. Among 69 individual drugs or drug classes with BBWs, we excluded beta-blocker medications from the analysis because we were unable to perform manual chart reviews that would have been required to fully assess the relevance of BBWs. Further, no alert was implemented for this class of drug during the intervention. Five other drugs were also excluded from the analysis as they were not used in the study clinics during the post-intervention period. Among the remaining 63 individual drugs or drug classes with BBWs, we analyzed the 30 more frequently prescribed medications that had BBW-related alerts implemented in the LMR during the intervention period. We calculated the proportion of medication prescription orders in which patients received a contraindicated drug, had a contraindicated disease, or did not receive adequate laboratory monitoring. We performed these calculations in both the pre- and post-intervention patient populations. We evaluated differences between the pre- and post-intervention periods using the chi-square test or Fisher exact test for discrete variables and the Student's *t*-test for continuous variables. We performed multiple logistic regression analyses of the outcome (violation of a BBW) applying the exchangeable covariance structure of generalized estimating equation (GEE) approach to adjust for within-patient correlations.[22] We examined potential risk factors as the independent variable, including the intervention (pre- vs. post-), and characteristics of the patients and providers. We included all variables with *p*-values less than 0.05 in the bivariate analyses in the baseline multivariable model. We examined interaction terms between covariates in the baseline model and retained them using a threshold *p*-value less than 0.05. We used backwards selection to drop insignificant effects one at a time. The baseline model included the following variables: intervention; patient age, sex, race, and language; provider type, site of care; number of medical problems, and medications. The model also included the following interaction terms: combinations of any two characteristic variables for patients at the beginning of the pre- and post-intervention period, combinations of any two characteristic variables for patients and providers when drugs with BBWs were prescribed, and combinations of intervention and every characteristic variable of patients and providers. Examples of interaction terms include: age group - × gender, provider type × number of medical problems on problem list, intervention × race, and intervention × site of clinic.

All variables in the final model were statistically significant ($p < 0.05$ with the score statistic in the Type 3 GEE analysis). We computed adjusted odds ratios (ORs) and 95% confidence intervals (CIs) based on the multiple logistic regression parameter estimates as measures of effect size. We performed all analyses using SAS System for Windows, version 8.2 (SAS Institute, Cary, NC).

RESULTS

During the pre-intervention period, 1014 providers prescribed 31 118 medication orders with BBWs to 24 477 patients. During the post-intervention period, 2270 providers prescribed 63 010 medication orders with BBWs to 45 744 patients.

The proportion of physicians who were primary care physicians was lower in the post-intervention period (48% vs. 36%, $p < 0.0001$), with more specialists (28% vs. 44%, $p < 0.0001$) post-intervention, because specialists were adopting the EHR during the study and our network focused on getting primary care providers on-line first. Among all 2661 providers, 623 (23%) functioned as providers in both periods. On average, each provider placed 31 medication orders with BBWs in the pre-intervention period and 28 medication orders with BBWs in the post-intervention period. Those 30 drugs/drug classes with BBW-related alerts implemented in the LMR

DOI: 10.1002/pds

accounted for 28 359 (91%) and 56 869 (90%) of all medication orders with BBWs in the pre- and post-intervention period, respectively.

Characteristics of the study patients and providers

A summary and comparison of patient demographic characteristics at the beginning of the pre- and post-intervention periods are shown in Table 1. Patients in the post-intervention period were more likely to be younger, male, white and without insurance of poverty relative to patients in the pre-intervention period. During the post-intervention period, specialists prescribed more orders with BBWs, while primary care physicians prescribed fewer orders with BBWs (Table 2). Providers in hospital-based clinics prescribed more orders with BBWs in the post-intervention period, while providers in other settings prescribed fewer. More medications with BBWs were prescribed to patients with fewer (<4) medical problems on the problem list and with more (>3) medications on their active medication list.

Table 1. Patient characteristics at the beginning of the pre- and post-intervention periods

Characteristic	Pre-intervention ($n = 23\,056$)	Post-intervention ($n = 42\,615$)	p
Age (years), No. (%)			0.0001
18–44	3448 (15.0)	6965 (16.3)	0.0001
45–54	4268 (18.5)	8303 (19.5)	0.003
55–64	5755 (25.0)	10 881 (25.5)	0.011
65–74	4772 (20.7)	8524 (20.0)	0.003
≥75	4813 (20.9)	7942 (18.6)	0.0001
Female, No. (%)	13 443 (58.3)	22 895 (53.7)	0.0001
Race/ethnicity, *n* (%)			0.0001
White	14 540 (63.1)	29 772 (69.9)	0.0001
Non-white	6198 (26.9)	9057 (21.3)	0.0001
Unknown	2318 (10.1)	3786 (8.9)	0.0001
Language, No. (%)			0.0001
English	19 363 (84.0)	35 382 (83.0)	0.002
Non-English	2844 (12.3)	4654 (10.9)	0.0001
Unknown	849 (3.7)	2579 (6.1)	0.0001
Insurance of poverty,* No. (%)			
Yes	3693 (16.0)	5276 (12.4)	0.0001

BBW, black-box warning.
*Medicaid or free care.

Non-adherence to BBWs in the pre- and post-intervention periods

Among all medication orders with alerts implemented during the intervention period (Table 3), there were slightly more violations of the BBW after the intervention (4.8% vs. 5.1%, $p = 0.045$). The frequency of orders with violations of BBWs was not significantly different between the pre- and post-intervention periods for drug–disease and drug–laboratory interactions. In contrast, there was a decrease in the frequency of BBW violations in the post-intervention period among orders with drug–drug and drug–pregnancy interactions (drug–drug interaction, 6.1% vs. 1.8%, $p < 0.0001$; drug–pregnancy interaction, 5.1% vs. 3.6%, $p = 0.01$).

The frequency of violation of BBWs for individual medications with alerts in the LMR is shown in Table 4. During the post-intervention period, a significant

Table 2. Characteristics of patients and providers when drugs with BBWs were prescribed

Characteristic	Pre-intervention ($n = 28\,359$)	Post-intervention ($n = 56\,869$)	p
Provider type, No. (%)			0.0001
Primary care MD, NP, or PA	21 954 (77.4)	41 589 (73.1)	0.0001
Specialist MD	3479 (12.3)	9195 (16.2)	0.0001
Other	2926 (10.3)	6085 (10.7)	0.09
Site of care, No. (%)			0.0001
Community-based private office	4589 (16.2)	8053 (14.2)	0.0001
Community health center	7674 (27.1)	12 091 (21.3)	0.0001
Hospital-based clinic	16 096 (56.8)	36 725 (64.6)	0.0001
Number of medical problems on problem list,* No. (%)			0.0001
0	8490 (29.9)	18 241 (32.1)	0.0001
1–3	5812 (20.5)	12 175 (21.4)	0.002
4–6	7029 (24.8)	14 000 (24.6)	0.59
≥7	7028 (24.8)	12 453 (21.9)	0.0001
Number of medications,† No. (%)			0.0001
0	8931 (31.5)	13 244 (23.3)	0.0001
1–3	9712 (34.3)	14 176 (24.9)	0.0001
≥4	9716 (34.3)	29 449 (51.8)	0.0001

BBW, black-box warning; MD, medical doctor; NP, nurse practitioner; PA, physician assistant.
*Defined as the number of medical problems on the problem list when the drug with BBWs was prescribed.
†Defined as the number of medications taken by the patient when or before the drug with BBWs was prescribed.

DOI: 10.1002/pds

Table 3. Frequencies of prescription orders non-adherent to the BBW

Type of BBW	Pre-intervention	Post-intervention	*p*
Drug–drug interaction, No. (%)	23/379 (6.1)	17/924 (1.8)	<0.0001
Drug–laboratory, No. (%)	1224/8538 (14.3)	2680/18 257 (14.7)	0.46
Drug–disease interaction, No. (%)	41/5292 (0.8)	97/10 233 (1.0)	0.28
Drug–pregnancy interaction, No. (%)	78/1517 (5.1)	120/3297 (3.6)	0.01
All, No. (%)	1357/28 359 (4.8)	2902/56 869 (5.1)	0.045

BBW, black-box warning.

decrease in the frequency of violation of the BBWs was found for clozapine, ticlopidine, ketorolac, metformin, isotretinoin, and ketoconazole while an increase was found for drugs including amiodarone, carbamazepine, and thioridazine. After excluding the level 3 and level 4 alerts, the violation rate decreased in the post-intervention period (0.46% vs. 0.32%, $p = 0.01$).

Multivariate correlates of BBW violation

In multivariate analyses, the final models included the following variables: intervention; patient age, sex, race, and language; provider type and site of care; and number of medical problems and medications (Table 5). Patients who were 75 years and older, had 0–3 medical problems, were not on any medication and who were seen by a primary care provider or a specialist at the hospital-based clinics were significantly more likely to receive a drug in violation of a BBW than were other patients (all $p \leq 0.03$). Patients in the pre-intervention period were less likely to be prescribed a drug in violation of a BBW when compared to those after the intervention period (OR 0.67, 95% CI, 0.47–0.96).

DISCUSSION

About 1 in 20 medication prescriptions written during the post-intervention period for drugs that had a black-box warning were in violation of a BBW. The implementation of BBW-related alerts as advanced decision support for electronic prescribing in the LMR had no impact on improvement of the overall BBW adherence rate, although we did observe modest improvement for specific subcategories which may be of particular clinical importance.

Computerized order entry (CPOE) systems with advanced decision support through alerts and reminders at the time of prescribing have been shown to be effective for reducing medication errors for inpatients, with a relative risk reduction for medication errors of 20–99%.[13,23–31] In comparison, relatively few studies have evaluated the effects of computerized decision support on medication errors in the outpatient setting. In one recent study to evaluate the effect of CPOE with decision support for selected medications prescribed in primary care clinics, Feldstein *et al.* [32] found that medication interaction alerts reduced warfarin-interacting medication prescription rates by 14.9% after the alert was implemented. The list of drugs with BBWs in our study did not include warfarin because the FDA did not begin warfarin-based BBWs until after our post-intervention period.[33] However, general alerts for warfarin–drug interactions and warfarin–pregnancy interactions were implemented in the LMR during the intervention period.

Why was there no improvement in adherence to BBWs in this study? It is possible that physicians may generally consider the warnings unimportant, since violations of BBWs were not found to result in increased levels of detectable harm in this patient population, though low levels of harm could have been missed.[18] Thus, alerting the physicians to the presence of a particular warning may make little difference to their adherence rates. We have previously found that high-level alerts—which in our system are level 1 and level 2 alerts—result in a high proportion of cancellations.[20] However, level 3 and level 4 alerts (indicating a lower level of potential harm) were common among the BBW medications and these appear to have little effect on behavior. We previously found that non-interruptive medication laboratory monitoring alerts did not improve receipt of recommended laboratory test monitoring in ambulatory care.[34] Regarding including all BBW alerts in CPOE, we would on the basis of these data recommend that other hospitals not include them, except for the clinically important subcategories which represent a small proportion of all BBWs.

In this study, we identified independent risk factors for prescribing in violation of BBWs. These included older patient age, and being seen by a primary care provider or a specialist at the hospital-based clinics. Patients not on any medication and having no or fewer

DOI: 10.1002/pds

Table 4. Frequencies of individual drug prescription orders non-adherent to the BBW

Drug name	Warning or monitoring recommendation*	Level of Alert related to BBW	Pre-intervention	Post-intervention	*p*-Value
Amiloride	Monitor serum potassium levels [every 6 months]	Level 4 (reminder 12 months) and level 3 (baseline 12 months)	32/197 (16.2)	65/284 (22.9)	0.074
Amiodarone	Monitor liver enzymes on high dose [every 6 months]	Level 4 (reminder 12 months) and level 3 (baseline 6 months)	16/263 (6.1)	120/745 (16.1)	<0.0001
ACE Inhibitors and Angiotensin II Antagonists	Unsafe in pregnancy	Level 2 alert if patient is pregnant	0/1424 (0)	0/3019 (0)	N/A
Bumetanide	Monitor electrolytes, creatinine, and BUN [every 6 months]	Level 3 baseline serum potassium only (12 months)	9/55 (16.4)	15/123 (12.2)	0.452
Carbamazepine	Monitor CBC-platelets at baseline and during therapy [once per year]	Level 3 baseline platelets only (6 months)	134/566 (23.7)	303/1031 (29.4)	0.014
Cisapride	Drug interactions with agents causing QT prolongation	DDI (levels 1–3)	0/6 (0)	0/5 (0)	N/A
Clozapine	(1) Do not initiate if WBC < 3500	Level 3 baseline WBC only (14 days)	0	0	
	(2) Do not initiate if history of myeloproliferative disease		0	0	
	(3) Weekly WBC for first 6 months of continuous treatment; if WBC ≥ 3000 may reduce to every other week; post-treatment WBC each week for 4 weeks		45/51 (88.2)	97/145 (66.9)	0.003
	Non-adherence to any of above		45/51 (88.2)	97/145 (66.9)	0.003
Cyclosporine	(1) Monitor renal function during therapy [every 3 months]	Level 3 (baseline 6 months) and level 4 (reminder 12 moths) for serum creatinine	36/136 (26.5)	81/394 (20.6)	0.15
	(2) If administered with methotrexate monitor CBC and LFTs monthly		1/136 (0.7)	4/394 (1.0)	1.00
	Non-adherence to any of above		36/136 (26.5)	83/394 (21.1)	0.193
Danazol	Negative pregnancy test prior to therapy	Level 2 alert if patient is pregnant and level 3 if patient is not pregnant but of child bearing potential	4/4 (100.0)	3/3 (100.0)	N/A
Ergotamine or dihydroergotamine	Concurrent use of CYP 3A4 inhibitors is contraindicated	DDI (levels 1–3)	7/56 (12.5)	10/37 (27.0)	0.076
Ganciclovir	Monitor CBC-platelets [every 2 weeks]	Level 3 (baseline 14 days) and level 4 (reminder 12 months) WBC only	38/61 (62.3)	100/207 (48.3)	0.055
Isoniazid	Not for use in patient with "active liver disease." [transaminases > 5 times upper limit of normal]	Level 3 (baseline 6 months) ALT only	0/629 (0)	0/996 (0)	N/A
Isotretinoin	Negative urine or serum pregnancy test (2 samples)	Category XX level 1 alert if patient is pregnant	15/24 (62.5)	40/147 (27.2)	0.0006
Itraconazole	(1) May not be administered to patients with onychomycosis and CHF	DDI (level 2 and level 3)			
	(2) Drug interactions with: cisapride, pimozide, quinidine, dofetilide				
	Non-adherence to any of above		0/91 (0)	0/112 (0)	N/A
Ketoconazole	(1) Monitor LFTs and bilirubin at baseline and then frequent intervals [3 weeks after baseline]	DDI (levels 1–3)	92/112 (82.1)	223/317 (70.4)	0.015
	(2) Drug interaction with cisapride, astemizole, terfenadine		0/112 (0)	0/317 (0)	N/A
	Non-adherence to any of above		92/112 (82.1)	223/317 (70.4)	0.015
Ketorolac	(1) Concurrent use with NSAIDs contraindicated	DDI (level 2) and level 2 critical serum creatinine	13/56 (23.2)	0/141 (0)	<0.0001
	(2) Contraindicated in patients with advanced renal impairment [creatinine ≥ 3]		0/56 (0)	1/141 (0.7)	1.00
	(3) Contraindicated in patients with peptic ulcers, GI bleeding and/or perforation (active or history)		1/56 (1.8)	2/141 (1.4)	1.00
	Non-adherence to any of above		14/56 (25.0)	3/141 (2.1)	<0.0001
Leflunomide	Exclude pregnancy prior to initiation of therapy	Category XX level 1 alert if pt is pregnant	53/54 (98.2)	65/67 (97.0)	1.00

DOI: 10.1002/pds

Metformin	(1) Avoid in patients with hepatic disease [transaminases > 3 times upper limit of normal]		14/4219 (0.3)	60/8810 (0.7)	0.01
	(2) Avoid in patients with renal dysfunction [Cr > 1.4 in women or > 1.5 in men]	Creatinine alert levels 2–4 (all 12 months) or CrCl 15 and 50 (Creatinine clearance alert)	61/4219 (1.5)	127/8810 (1.4)	0.98
	(3) Patients ≥ age 80 need creatinine clearance measured		123/4219 (2.9)	129/8810 (1.5)	<0.0001
	(4) Avoid in patients with CHF requiring pharmacologic treatment		25/4219 (0.6)	33/8810 (0.4)	0.08
	Non-adherence to any of above		216/4219 (5.1)	344/8810 (3.9)	0.001
Misoprostol	Women who take drug for NSAID ulcer reduction should have negative HCG	Category XX level 1 alert if pt is pregnant	0/5 (0)	0/22 (0)	N/A
Nefazodone	Contraindicated in patients with 'active liver disease' [AST or ALT ≥ 3 times upper limit of normal], Withdraw therapy if AST or ALT ≥ 3 times upper limit of normal	ALT level 2 (12 months) and 3 (6 months)	0/283 (0)	0/140 (0)	N/A
Nevirapine	Patients should be closely monitored at baseline and for first 12 weeks for signs of liver reactions [monitor LFTs every 4 months]	ALT level 3 (baseline 12 months) only	21/78 (26.9)	9/63 (14.3)	0.068
Procainamide	CBC-diff-platelets weekly during first 3 months and periodically thereafter [once per year]	HCT and platelet level 3 (baseline 6 months) and level 4 (reminder 12 months)	0/5 (0)	0/5 (0)	N/A
Ribavirin/Interferon	Contraindicated in pregnancy	Category X level 2 alert if patient is pregnant	0/0 (0)	0/0 (0)	N/A
Ritonavir	Drug interactions with some non-sedating antihistamines, sedative hypnotics, antiarrhythmics, and ergot alkaloids	DDI (levels 1–3)	0/44 (0)	0/278 (0)	N/A
Thalidomide	Contraindicated in pregnancy, negative HCG 24 hours prior to therapy initiation, weekly during first month, monthly thereafter	Category XX level 1 alert if patient is pregnant	6/6 (100)	12/21 (57.1)	0.07
Thioridazine	(1) Contraindicated with P450 2D6 inhibitors and agents that prolong QT	DDI (level 2 and level 3)	3/14 (21.4)	7/34 (20.6)	0.3
	(2) Contraindicated in patients with history of cardiac arrhythmia or congenital long QT syndrome		0/14 (0)	1/34 (2.9)	1.00
	(3) Monitor serum potassium [once per year]		7/14 (50.0)	33/34 (97.1)	0.0003
	(4) Baseline and periodic EKG [once per year]		10/14 (71.4)	29/34 (85.3)	0.42
	Non-adherence to any of above		11/14 (78.6)	34/34 (100)	0.02
Ticlopidine	CBC-diff, platelets, and smear at baseline and every 2 weeks during first 3 months. If patient stops therapy in first 3 months continue monitoring for 2 more weeks.	HGB and neutrophils level 3 (baseline 12 months)	12/12 (100)	11/18 (61.1)	0.024
Tretinoin	Negative serum or urine pregnancy test within 1 week prior to therapy	Category D level 2 alert if pt is pregnant	0/0 (0)	0/18 (0)	N/A
Triamterene	Monitor serum potassium [every 6 months]	Serum potassium levels 2–4 (12 months)	389/1785 (21.8)	902/4195 (21.5)	0.8
Valproate	Baseline LFTs and frequent monitoring [once per year]	ALT level 3 (baseline 12 months)	207/701 (29.5)	463/1746 (26.5)	0.12

ACE, angiotensin-converting enzyme; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BBW, black-box warning; BUN, blood urea nitrogen; CBC, complete blood cell; CHF, congestive heart failure; Cr, creatinine; CrCl, creatinine clearance; DDI, drug drug interaction; EKG, electrocardiography; GI, gastrointestinal; HGB, hemoglobin; HCG, human chorionic gonadotropin; LFT, liver function test; NSAID, non-steroidal anti-inflammatory drug; N/A, non-applicable; WBC, white blood cell.

*In cases where the black box warning was unclear, an operational definition appears in brackets, as determined by discussions with specialists.

DOI: 10.1002/pds

Table 5. Multivariate analysis for risk of non-adherence to the BBW

Variables	OR (95% CI)*
Intervention	
Pre-	0.67 (0.47–0.96)
Post-	1.00
Age, years	
18–44	0.60 (0.39–0.92)
45–54	0.36 (0.23–0.57)
55–64	0.44 (0.28–0.70)
65–74	0.36 (0.22–0.60)
≥75	1.00
Sex	
Female	1.13 (0.96–1.33)
Male	1.00
Race/ethnicity	
White	1.27 (0.88–1.83)
Non-white	1.00
Unknown	1.03 (0.58–1.82)
Language	
English	0.77 (0.57–1.04)
Non-English	1.00
Unknown	0.57 (0.36–0.92)
Insurance of poverty	
Yes	1.00
No	1.12 (0.93–1.33)
Provider type	
Primary care MD, NP, or PA	1.61 (1.04–2.48)
Specialist MD	1.80 (1.09–2.98)
Other	1.00
Site of care	
Community-based private office	1.00
Community health center	1.50 (0.90–2.50)
Hospital-based clinic	2.15 (1.41–3.26)
Number of medical problems on problem list†	
0	1.43 (1.14–1.80)
1–3	1.40 (1.12–1.74)
4–6	1.23 (0.98–1.54)
≥7	1.00
Number of medications‡	
0	2.11 (1.16–3.84)
1–3	1.00
≥4	1.34 (0.86–2.07)

BBW, black-box warning; OR, odds ratio; CI, 95% confidence interval; MD, medical doctor; NP, nurse practitioner; PA, physician assistant.

*Final model including 10 single terms and 12 interaction terms: age group × (gender, race, language), race × insurance, provider type × (site of clinic, number of medications), number of medications × (site of clinic, number of problems), intervention × (race, language, site of clinic, number of medication); All variables with $p < 0.05$ in final model.

†Defined as the number of medical problems on the problem list when the drug with BBWs was prescribed.

‡Defined as the number of medications taken by the patient when or before the drug with BBWs was prescribed.

medical problems were also at higher risk of overrides. Other studies have found little clustering.[35]

This study has a number of implications for the FDA warning process. Adapting this set of warnings for clinical decision support made it clear that the BBWs would be more useful and easier to implement in computerized systems generally if the need for implementation in clinical decision support was explicitly considered at the time of development. In addition, many of the warnings appear to be of marginal value given the frequency with which they are overridden. One function of the Sentinel Network which has been sponsored by the FDA might be to evaluate how specific warnings are regarded by prescribers, whether or not they are overridden, and if they are overridden, what is the absolute risk of associated harm. Some warnings could be rescinded if it appears the risk is low.

Our study has a number of limitations. We did not assess the frequency of harm associated with alert overrides in the current study. In addition, the study was done as a before–after trial, because it was judged that it would have been unethical to withhold black-box warnings, so that we did not have a contemporaneous control group. Having a contemporaneous control group would have been a stronger design. For medications requiring laboratory testing at baseline and/or during therapy, we determined the adherence/violation status based on the reports of laboratory results and medication start and stop dates. We may have overestimated the frequency of prescribing violation as laboratory monitoring tests could be done outside the study clinics. In addition, providers could tell patients to hold a medication in response to a specific circumstance without discontinuing the drug from the patient's medication list. Finally, this study was conducted in urban medical practices affiliated with academic teaching centers so that the results may not be generalizable to other settings.

We evaluated the effectiveness of implementing black-box related alerts in an EHR in the outpatient setting. Our results suggest that current EHR alerts did not improve overall adherence to the BBWs for drug–drug, drug–laboratory, drug–disease, and drug–pregnancy interactions, although we observed improvement in sub-group comparisons for drug–drug and drug–pregnancy interactions, and for some individual drugs. Further evaluation is needed regarding the absolute risk associated with specific BBWs.

KEY POINTS

- Implementation of alerts about black-box warnings through electronic prescribing in outpatients had no impact on improvement of the overall black-box warnings adherence rate.
- However, alerts did improve adherence for a few specific clinically important subcategories.

DOI: 10.1002/pds

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by grant from Partners Information Systems Research Council, Partners HealthCare System, Boston, MA. The funding sponsor had no role in study design, conduct, analysis, interpretation of the data, or preparation, review or approval of the manuscript. The authors thank Hanna M. Seidling, PhD, and Jennifer Boehne, PharmD for their comments on draft of this article.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix. List of drugs with a Black Box Warning pertaining to drug-drug interactions, drug-laboratory interactions, drug-disease interactions and/or drug-pregnancy interactions.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

REFERENCES

1. Classen DCM, Pestotnik SLM, Evans RSP, Lloyd JF, Burke JPM. Adverse drug events in hospitalized patients: Excess length of stay, extra costs, and attributable mortality. *JAMA* 1997; **277**(4): 301–306.
2. Bates DW, Cullen DJ, Laird N, *et al*. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. *JAMA* 1995; **274**(1): 29–34.
3. Jha AK, Kuperman GJ, Teich JM, *et al*. Identifying adverse drug events: Development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *J Am Med Inform Assoc* 1998; **5**(3): 305–314.
4. Bates DW, Boyle DL, Vander Vliet MB, Schneider J, Leape L. Relationship between medication errors and adverse drug events. *J Gen Intern Med* 1995; **10**(4): 199–205.
5. Hug BL, Witkowski DJ, Sox CM, *et al*. Adverse drug event rates in six community hospitals and the potential impact of computerized physician order entry for prevention. *J Gen Intern Med* 2010; **25**: 31–38.
6. Gandhi TK, Weingart SN, Borus J, *et al*. Adverse drug events in ambulatory care. *N Engl J Med* 2003; **348**(16): 1556–1564.
7. Institute of Medicine. *Preventing Medication Errors: Quality Chasm Series*. National Academies Press: Washington, DC, 2006.
8. Leape LL, Bates DW, Cullen DJ, *et al*. Systems analysis of adverse drug events. ADE Prevention Study Group. *JAMA* 1995; **274**(1): 35–43.
9. Shamliyan TA, Duval S, Du J, Kane RL. Just what the doctor ordered. Review of the evidence of the impact of computerized physician order entry system on medication errors. *Health Serv Res* 2008; **43**: 32–53.
10. Wolinsky JI, Gurwitz JH, Field TS, *et al*. The effect of computerized physician order entry with clinical decision support on the rates of adverse drug events: A systematic review. *J Gen Intern Med* 2008; **23**: 451–458.
11. Gurwitz JH, Field TS, Harrold LR, *et al*. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. *JAMA* 2003; **289**(9): 1107–1116.
12. Lazarou JM, Pomeranz BHM, Corey PNP. Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. *JAMA* 1998; **279**(15): 1200–1205.
13. Bates DW, Teich JM, Lee J, *et al*. The impact of computerized physician order entry on medication error prevention. *J Am Med Inform Assoc* 1999; **6**(4): 313–321.
14. King WJ, Paice N, Rangrej J, Forestell GJ, Swartz R. The effect of computerized physician order entry on medication errors and adverse drug events in pediatric inpatients. *Pediatrics* 2003; **112**: 506–509.
15. Gandhi TK, Weingart SN, Seger AC, *et al*. Outpatient prescribing errors and the impact of computerized prescribing. *J Gen Intern Med* 2005; **20**(9): 837–841.
16. The Value of Computerized Provider Order Entry in Ambulatory Settings. Available at: www.citl.org/research/ACPOE_Executive_preview.pdf [accessed from 8 January 2006].
17. Murphy SM, Roberts RM. ''Black box'' 101: How the Food and Drug Administration evaluates, communicates, and manages drug benefit/risk. *J Allergy Clin Immunol* 2006; **117**(1): 34–39.
18. Lasser KE, Seger DL, Yu DT, *et al*. Adherence to black box warnings for prescription medications in outpatients. *Arch Intern Med* 2006; **166**(3): 338–344.
19. Wagner AK, Chan KA, Dashevsky I, *et al*. FDA drug prescribing warnings: Is the black box half empty or half full? *Pharmacoepidemiol Drug Saf* 2006; **15**: 369–386.
20. Shah NR, Seger AC, Seger DL, *et al*. Improving acceptance of computerized prescribing alerts in ambulatory care. *J Am Med Inform Assoc* 2006; **13**(1): 5–11.
21. Paterno MD, Maviglia SM, Gorman PN, *et al*. Tiering drug–drug interaction alerts by severity increases compliance rates. *J Am Med Inform Assoc* 2009; **16**(1): 40–46.
22. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986; **42**: 121–130.
23. Ammenwerth E, Schnell-Inderst P, Machan C, Siebert U. The effect of electronic prescribing on medication errors and adverse drug events: A systematic review. *J Am Med Inform Assoc* 2008; **15**(5): 585–600.
24. Potts AL, Barr FE, Gregory DF, Wright L, Patel NR. Computerized physician order entry and medication errors in a pediatric critical care unit. *Pediatrics* 2004; **113**(1): 59–63.
25. Pestotnik SL, Classen DC, Evans RS, Burke JP. Implementing antibiotic practice guidelines through computer-assisted decision support: Clinical and financial outcomes. *Ann Intern Med* 1996; **124**(10): 884–890.
26. Cordero L, Kuehn L, Kumar RR, Mekhjian HS. Impact of computerized physician order entry on clinical practice in a newborn intensive care unit. *J Perinatol* 2004; **24**(2): 88–93.
27. Colpaert K, Claus B, Somers A, Vandewoude K, Robays H, Decruyenaere J. Impact of computerized physician order entry on medication prescription errors in the intensive care unit: A controlled cross- sectional trial. *Crit Care* 2006; **10**(1): R21.
28. Kim GR, Chen AR, Arceci RJ, *et al*. Error reduction in pediatric chemotherapy: Computerized order entry and failure modes and effects analysis. *Arch Pediatr Adolesc Med* 2006; **160**(5): 495–498.
29. Evans RS, Pestotnik SL, Classen DC, *et al*. A computer-assisted management program for antibiotics and other antiinfective agents. *N Engl J Med* 1998; **338**: 232–238.
30. Oliven A, Michalake I, Zalman D, Dorman E, Yeshurun D, Odeh M. Prevention of prescription errors by computerized, on-line surveillance of drug order entry. *Int J Med Inform* 2005; **74**(5): 377–386.
31. Tamblyn R, Huang A, Perreault R, *et al*. The medical office of the 21st century (MOXXI): Effectiveness of computerized decision-making support in reducing inappropriate prescribing in primary care. *Can Med Assoc J* 2003; **169**(6): 549–556.
32. Feldstein AC, Smith DH, Perrin N, *et al*. Reducing warfarin medication interactions: An interrupted time series evaluation. *Arch Intern Med* 2006; **166**(9): 1009–1015.
33. US Food and Drug Administration. MedWatch Web site. FDA and Bristol-Myers Squibb notified pharmacists and physicians of revisions to the labeling for Coumadin. Available at: http://www.fda.gov/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm150518.htm [accessed from 25 Oct 2009].
34. Lo HG, Matheny ME, Seger DL, Bates DW, Gandhi TK. Impact of non-interruptive medication laboratory monitoring alerts in ambulatory care. *J Am Med Inform Assoc* 2009; **16**(1): 66–71.
35. Bates DW, Miller EB, Cullen DJ, *et al*. Patient risk factors for adverse drug events in hospitalized patients. ADE Prevention Study Group. *Arch Intern Med* 1999; **159**(21): 2553–2560.

OPEN ACCESS Freely available online

PLoS GENETICS

Eight Common Genetic Variants Associated with Serum DHEAS Levels Suggest a Key Role in Ageing Mechanisms

Guangju Zhai[1☯], Alexander Teumer[2☯], Lisette Stolk[3,4☯], John R. B. Perry[5,6☯], Liesbeth Vandenput[7☯], Andrea D. Coviello[8☯], Annemarie Koster[9☯], Jordana T. Bell[1,6], Shalender Bhasin[10], Joel Eriksson[7], Anna Eriksson[7], Florian Ernst[2], Luigi Ferrucci[11], Timothy M. Frayling[5], Daniel Glass[1], Elin Grundberg[1,12], Robin Haring[13], Åsa K. Hedman[6], Albert Hofman[4,14], Douglas P. Kiel[15], Heyo K. Kroemer[16], Yongmei Liu[17], Kathryn L. Lunetta[18], Marcello Maggio[19], Mattias Lorentzon[7], Massimo Mangino[1], David Melzer[5], Iva Miljkovic[20], MuTHER Consortium, Alexandra Nica[12,21], Brenda W. J. H. Penninx[22], Ramachandran S. Vasan[8,23], Fernando Rivadeneira[3,4], Kerrin S. Small[1,12], Nicole Soranzo[1,12], André G. Uitterlinden[3,4,14], Henry Völzke[24], Scott G. Wilson[1,25,26], Li Xi[27], Wei Vivian Zhuang[18], Tamara B. Harris[9¶], Joanne M. Murabito[8,23¶], Claes Ohlsson[7¶], Anna Murray[5¶], Frank H. de Jong[3¶], Tim D. Spector[1]*, Henri Wallaschofski[13¶]

1 Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom, **2** Interfaculty Institute for Genetics and Functional Genomics, University of Greifswald, Greifswald, Germany, **3** Department of Internal Medicine, Erasmus MC Rotterdam, Rotterdam, The Netherlands, **4** Netherlands Consortium of Healthy Ageing, Rotterdam, The Netherlands, **5** Genetics of Complex Traits, Peninsula Medical School, University of Exeter, Exeter, United Kingdom, **6** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **7** Department of Internal Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, **8** Sections of General Internal Medicine, Preventive Medicine, Cardiology and Endocrinology, Diabetes and Nutrition, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America, **9** Laboratory for Epidemiology, Demography, and Biometry, National Institute on Aging, Bethesda, Maryland, United States of America, **10** Section of Endocrinology, Diabetes, and Nutrition, Claude D. Pepper Older Americans Independence Center, Boston University School of Medicine, Boston, Massachusetts, United States of America, **11** Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, United States of America, **12** Wellcome Trust Sanger Institute, Hixton, United Kingdom, **13** Institute for Clinical Chemistry and Laboratory Medicine, University of Greifswald, Greifswald, Germany, **14** Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands, **15** Hebrew Senior Life Institute for Aging Research and Harvard Medical School, Boston, Massachusetts, United States of America, **16** Center of Pharmacology and Experimental Therapeutics, Department of Pharmacology, University of Greifswald, Greifswald, Germany, **17** Department of Epidemiology and Prevention, Wake Forest University Health Sciences, Winston-Salem, North Carolina, United States of America, **18** Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, **19** Department of Internal Medicine and Biomedical Sciences, Section of Geriatrics, University of Parma, Parma, Italy, **20** Department of Epidemiology, University of Pittsburgh, Pittsburg, Pennsylvania, United States of America, **21** Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland, **22** Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands, **23** The National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, United States of America, **24** Institute for Community Medicine, University of Greifswald, Greifswald, Germany, **25** Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands, Australia, **26** School of Medicine and Pharmacology, University of Western Australia, Nedlands, Australia, **27** Molecular Medicine – Computational Biology, Pfizer Worldwide R&D, Groton, Connecticut, United States of America

Abstract

Dehydroepiandrosterone sulphate (DHEAS) is the most abundant circulating steroid secreted by adrenal glands—yet its function is unknown. Its serum concentration declines significantly with increasing age, which has led to speculation that a relative DHEAS deficiency may contribute to the development of common age-related diseases or diminished longevity. We conducted a meta-analysis of genome-wide association data with 14,846 individuals and identified eight independent common SNPs associated with serum DHEAS concentrations. Genes at or near the identified loci include *ZKSCAN5* (rs11761528; $p = 3.15 \times 10^{-36}$), *SULT2A1* (rs2637125; $p = 2.61 \times 10^{-19}$), *ARPC1A* (rs740160; $p = 1.56 \times 10^{-16}$), *TRIM4* (rs17277546; $p = 4.50 \times 10^{-11}$), *BMF* (rs7181230; $p = 5.44 \times 10^{-11}$), *HHEX* (rs2497306; $p = 4.64 \times 10^{-9}$), *BCL2L11* (rs6738028; $p = 1.72 \times 10^{-8}$), and *CYP2C9* (rs2185570; $p = 2.29 \times 10^{-8}$). These genes are associated with type 2 diabetes, lymphoma, actin filament assembly, drug and xenobiotic metabolism, and zinc finger proteins. Several SNPs were associated with changes in gene expression levels, and the related genes are connected to biological pathways linking DHEAS with ageing. This study provides much needed insight into the function of DHEAS.

Citation: Zhai G, Teumer A, Stolk L, Perry JRB, Vandenput L, et al. (2011) Eight Common Genetic Variants Associated with Serum DHEAS Levels Suggest a Key Role in Ageing Mechanisms. PLoS Genet 7(4): e1002025. doi:10.1371/journal.pgen.1002025

Editor: Greg Gibson, Georgia Institute of Technology, United States of America

Received November 18, 2010; **Accepted** January 27, 2011; **Published** April 14, 2011

Funding: This is a collaborative work and supported by many sources of funding. Due to the limit of words here, the details and list of the funding sources are provided in the Acknowledgments. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tim.spector@kcl.ac.uk

☯ These authors contributed equally to this work.

¶ These senior authors were joint senior authors on this work.

Author Summary

Dehydroepiandrosterone sulphate (DHEAS), mainly secreted by the adrenal gland, is the most abundant circulating steroid in humans. It shows a significant physiological decline after the age of 25 and diminishes about 95% by the age of 85 years, which has led to speculation that a relative DHEAS deficiency may contribute to the development of common age-related diseases or diminished longevity. Twin- and family-based studies have shown that there is a substantial genetic effect with heritability estimate of 60%, but no specific genes regulating serum DHEAS concentration have been identified to date. Here we take advantage of recent technical and methodological advances to examine the effects of common genetic variants on serum DHEAS concentrations. By examining 14,846 Caucasian individuals, we show that eight common genetic variants are associated with serum DHEAS concentrations. Genes at or near these genetic variants include *BCL2L11*, *ARPC1A*, *ZKSCAN5*, *TRIM4*, *HHEX*, *CYP2C9*, *BMF*, and *SULT2A1*. These genes have various associations with steroid hormone metabolism—co-morbidities of ageing including type 2 diabetes, lymphoma, actin filament assembly, drug and xenobiotic metabolism, and zinc finger proteins—suggesting a wider functional role for DHEAS than previously thought.

Introduction

Dehydroepiandrosterone sulphate (DHEAS), mainly secreted by the adrenal gland, is the most abundant circulating steroid in humans. It acts as an inactive precursor which is converted initially into DHEA and thereafter into active androgens and estrogens in peripheral target tissues [1]. In humans the serum concentration of circulating DHEAS is 100- to 500-fold or 1000 to 10,000 higher than that of testosterone and estradiol respectively. Unlike DHEA, which is swiftly cleared from the circulation and shows diurnal variation, serum DHEAS concentrations are stable and facilitate accurate measurement and diagnosis of pathology [2].

DHEAS is distinct from the other major adrenal steroids (cortisol and aldosterone) in showing a significant physiological decline after the age of 25 and diminishes about 95% by the age of 85 years [3]. This age-related decline has led to speculation that a relative DHEAS deficiency may contribute to the development of common age-related diseases or diminished longevity [4,5]. Low DHEAS concentrations are possibly associated with increased insulin resistance [6,7] and hypertension [8], but not with incident metabolic syndrome [9]. It is strongly associated with osteoporosis in women [10,11] but not in men [12]. Concurrent change in DHEAS tracks with declines in gait speed, modified mini-mental state examination score (3MSE), and digit symbol substitution test (DSST) in very old women but not in men [13]. Low circulating DHEAS is also strongly associated with cardiovascular disease and mortality in men [14–18] but not in women [19]. A recent 15-year follow-up study showed that DHEAS was negatively related to all-cause, all cancers, and other medical mortality, whereas high DHEAS concentrations were protective [20]. This has led to its widespread and uncontrolled use as a controversial anti-ageing and sexual performance supplement in the USA and other western countries without any clear data about efficacy, potential risks or benefits [21].

Despite these observations, the physiological function of DHEAS and its importance in maintaining health are poorly understood. Although previous twin [22,23] and family-based studies [24,25] have shown that there is a substantial genetic effect with a heritability estimate of 60% [22], no specific genes regulating serum DHEAS concentration in healthy individuals have been identified to date. Therefore, the current study meta-analyzed the results of genome-wide association studies (GWAS) performed in a total of 14,846 individuals from seven cohorts to identify common genetic variants associated with serum DHEAS concentrations. The findings not only advance understanding of how serum DHEAS concentration is regulated by genes but also provide clues as to its mechanism of action as well as Mendelian randomisation principles [26].

Results

We carried out a meta-analysis of 8,565 women and 6,281 men of European origin from collaborating studies: TwinsUK (n = 4,906), Framingham Heart Study (FHS) (n = 3,183), SHIP (n = 1,832), Rotterdam Study (RS1) (n = 1,597), InCHIANTI (n = 1,182), Health ABC (n = 1,222), and GOOD (n = 924). Serum samples were collected either after overnight fasting or non-fasting in each cohort and DHEAS was measured by either immunoassay or liquid chromatography tandem mass spectrometry (LC-MS/MS) methods (Table 1). Mean age differed across the cohorts from 19 to 74 years in men and 50 to 74 years in women and corresponding mean DHEAS concentrations varied from 1.20 to 7.05 μmol/L (Table 1).

Each cohort performed GWA tests for log transformed DHEAS on ~2.5 million imputed single nucleotide polymorphisms (SNPs) in men and women separately with adjustment for age, and additionally for age and sex for those cohorts who had data in both men and women. Then Z-scores from each cohort were pooled for the meta-analysis at each SNP.

In all our individual GWAS, λ_{GC}, which is defined as the median χ^2 (1 degree of freedom) association statistic across SNPs divided by its theoretical median under the null distribution [27], ranged from 0.984 to 1.023, indicating that there was no population stratification or it was very minor. Further, we corrected for population stratification by applying the genomic control method [27]; the λ_{GC} in the meta-analysis is 1.017. In addition, the effect direction was consistent across all the cohorts and there is no between-study heterogeneity as indicated by I^2 ranging between 0 and 0.12 (Table 2).

We found 44 SNPs were associated with serum DHEAS concentrations in men at conventional genome-wide significance ($p<5\times10^{-8}$), which are all located on chromosome 7q22.1 (Figure 1B; Table S1). All these SNPs except for three were significant in women (Figure 1A; Table S1). In addition, 19 SNPs located on chromosome 19q13.3 were found in women to be associated with serum DHEAS concentrations with $p<5\times10^{-8}$. In the sex-combined meta-analysis, the significance became stronger for all these SNPs (Figure 1C; Table S1). Further, we found 8 SNPs located on chromosome 10q23.33 which represents two regions more than 2 MB apart, 12 SNPs on chromosome 15q15.1, and in addition, 4 SNPs on chromosome 19q13.3 were associated with serum DHEAS concentrations with $p<5\times10^{-8}$. Together we found a total of 87 SNPs associated with serum DHEAS concentrations with $p<5\times10^{-8}$, representing five chromosomal regions of less than 1 Mb each (Table S1).

The most significantly associated SNPs in each of these five regions are presented in Table 2. The minor allele of rs11761528 ($p=3.15\times10^{-36}$) on chromosome 7q22.1, rs2637125 ($p=2.61\times10^{-19}$) on chromosome 19q13.3, and rs2497306 ($p=4.6\times10^{-9}$) and rs2185570 ($p=2.29\times10^{-8}$) on chromosome 10q22.33 (more than 2 Mb apart), were negatively associated with

Table 1. Descriptive statistics of serum levels of DHEAS (μmol/L) for each cohort.

Males									
Cohort	**Assay**	**Mean Age (Range)**	**Mean**	**SD**	**Median**	**Min**	**Max**	**Range**	**n**
RS1	Immunoassay	69 (55–98)	4.34	2.88	3.70	0.10	23.08	22.98	*740*
SHIP	Immunoassay	51 (20–79)	1.90	1.21	1.64	0.31	8.90	8.59	1832
FHS	Immunoassay	51 (25–80)	7.05	5.07	5.35	0.27	29.86	29.59	*1571*
GOOD	MassSpec	19 (18–20)	6.31	2.33	6.04	1.27	15.10	13.83	924
InCHIANTI	Immunoassay	67 (23–94)	3.16	2.98	2.25	0.02	33.06	33.04	*518*
HABC	Immunoassay	74 (69–80)	1.58	1.12	1.40	0.00	9.93	9.93	696
								n Total	*6281*
Females									
Cohort	**Assay**	**Mean Age (Range)**	**Mean**	**SD**	**Median**	**Min**	**Max**	**Range**	**n**
TwinsUK I	Immunoassay	50 (17–82)	3.95	2.47	3.40	0.20	19.30	19.10	2541
TwinsUK II	Immunoassay	50 (16–82)	4.21	2.79	3.60	0.10	22.30	22.20	*2365*
RS1	Immunoassay	69 (55–98)	2.65	2.03	2.11	0.01	13.61	13.60	*857*
FHS	Immunoassay	51 (22–77)	3.84	2.96	3.02	0.30	21.01	20.71	1612
InCHIANTI	Immunoassay	68 (21–95)	2.69	2.35	1.96	0.04	15.29	15.25	*664*
HABC	Immunoassay	74 (69–80)	1.20	0.88	0.97	0.00	5.59	5.59	526
								n Total	*8565*

doi:10.1371/journal.pgen.1002025.t001

DHEAS concentrations. In comparison, the minor allele of rs7181230 ($p = 5.44\times10^{-11}$) on chromosome 15q15.1 was positively associated with serum DHEAS concentrations. Based on the HapMap3 release2 CEU data, the significant 87 SNPs from within the five regions have low pair-wise r^2, indicating potentially multiple independent signals. To verify this, we performed a conditional meta-analysis with adjustment for the five most significant SNPs plus age and sex in each cohort.

After this adjustment, all other SNPs on chromosome 10, 15, and 19 became non-significant (Figure 1D). However, on chromosome 7, we found two independent signals; one defined by rs11761528 and a second located 370 kb upstream in the 3′ UTR of the *TRIM4* and *CYP3A43* genes (rs17277546, $p = 4.50\times10^{-11}$). Furthermore, we identified two additional significant loci associated with DHEAS, one on chromosome 2q13 (rs6738028, $p = 1.72\times10^{-8}$), and another on chromosome 7 within the *ARPC1A* gene (rs740160 located 161 kb downstream of rs11761528, $p = 1.56\times10^{-16}$) (Table 2; Figure 1D). In total, we found eight independent SNPs associated with serum DHEAS concentrations at conventional genome-wide significant level ($p<5\times10^{-8}$) (Table 2). The effect was consistently in the same direction across all cohorts (Table 2). No heterogeneity among cohorts was observed (Table 2). These SNPs together explained ~4% of the total and ~7% of genetic variance of serum DHEAS concentrations (based on TwinsUK data). To further look at whether the magnitude of these genetic association varies with age,

Table 2. SNPs associated with serum DHEAS concentrations: genome-wide results of meta-analysis of men and women combined.

SNP	**Chr**	**Position in base pair**	**Freq**	**Effect Allele**	**Beta (SE)***	**P value**	**I^2 index¶**	**Effect direction in each study**	**Gene**	**Distance to the gene**
Discovery meta-analysis										
rs11761528	7	98956737	0.08	T	−0.16 (0.01)	3.15×10^{-36}	0.12	−−−−−−−−	*ZKSCAN5*	intron
rs2637125	19	53093705	0.15	A	−0.09(0.01)	2.61×10^{-19}	0.00	−−−−−−−−	*SULT2A1*	12 kb
rs7181230	15	38148033	0.33	G	0.05(0.01)	5.44×10^{-11}	0.00	++++++++	*BMF*	23 kb
rs2497306	10	94475191	0.49	C	−0.04(0.01)	4.64×10^{-9}	0.00	−−−−−−−−	*HHEX*	25 kb
rs2185570	10	96741260	0.13	C	−0.06(0.01)	2.29×10^{-8}	0.00	−−−−−−−−	*CYP2C9*	−2 kb
Conditional analysis										
rs740160§	7	98795816	0.05	T	0.15 (0.02)	1.56×10^{-16}	0.02	++++++++	*ARPC1A*	intron
rs17277546§	7	99327507	0.05	A	−0.11 (0.02)	4.50×10^{-11}	0.00	−−−−−−−−	*TRIM4;CYP3A43*	3′UTR
rs6738028§	2	111665798	0.40	G	−0.04 (0.01)	1.72×10^{-8}	0.00	−−−−−−−−	*BCL2L11*	−62 kb

*Beta was expressed as natural log changes in serum DHEAS concentration in μmol/L per copy of the risk allele.
¶index for between-study heterogeneity: 0.25 – low, 0.50 – moderate and 0.75 – high heterogeneity.
§pre-conditional p values were 0.612, 1.90×10^{-26}, and 1.94×10^{-7} for rs740160, rs17277546, and rs6738028, respectively.
doi:10.1371/journal.pgen.1002025.t002

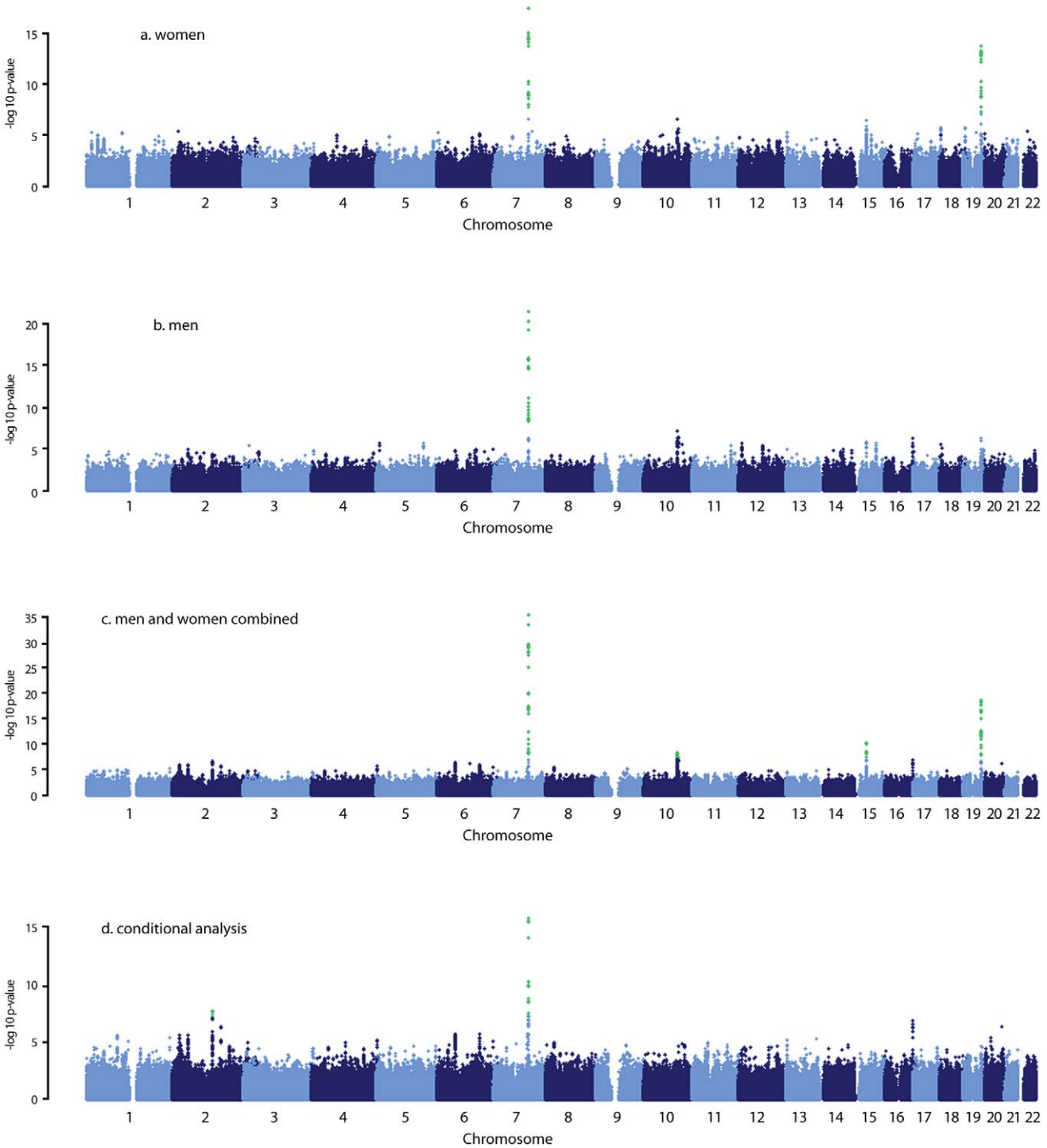


Figure 1. Manhattan plots for the genome-wide meta-analysis results. Green dots indicate the SNPs with $p<5\times10^{-8}$.
doi:10.1371/journal.pgen.1002025.g001

we carried out an interaction analysis between age and each of these 8 SNPs on serum DHEAS concentrations by including an interaction term of age×SNP in the linear regression model in each cohort and then meta-analyzed the results. We found that there was no significant interaction between age and each of these SNPs (all p values≥0.05).

The genes at, or near the identified SNPs, include *BCL2L11* on chromosome 2, *ZKSCAN5*, *ARPC1A*, *TRIM4* and *CYP3A43* on chromosome 7, *HHEX* and *CYP2C9* on chromosome 10, *BMF* on chromosome 15, and *SULT2A1* on chromosome 19 (Figure 2). To explore the potentially functional impacts and likely genetic mechanisms, we used two resources: Genome-wide expression data from the Multiple Tissue Human Expression Resource (MuTHER) [28] (http://www.muther.ac.uk/) based on ~777 unselected UK twins sampled for skin, adipose tissue, and lymphoblastoid cell lines (LCLs) (more details in Text S1); and published gene expression data in human liver [29]. We found that 3 DHEAS-associated SNPs were clearly associated with the related

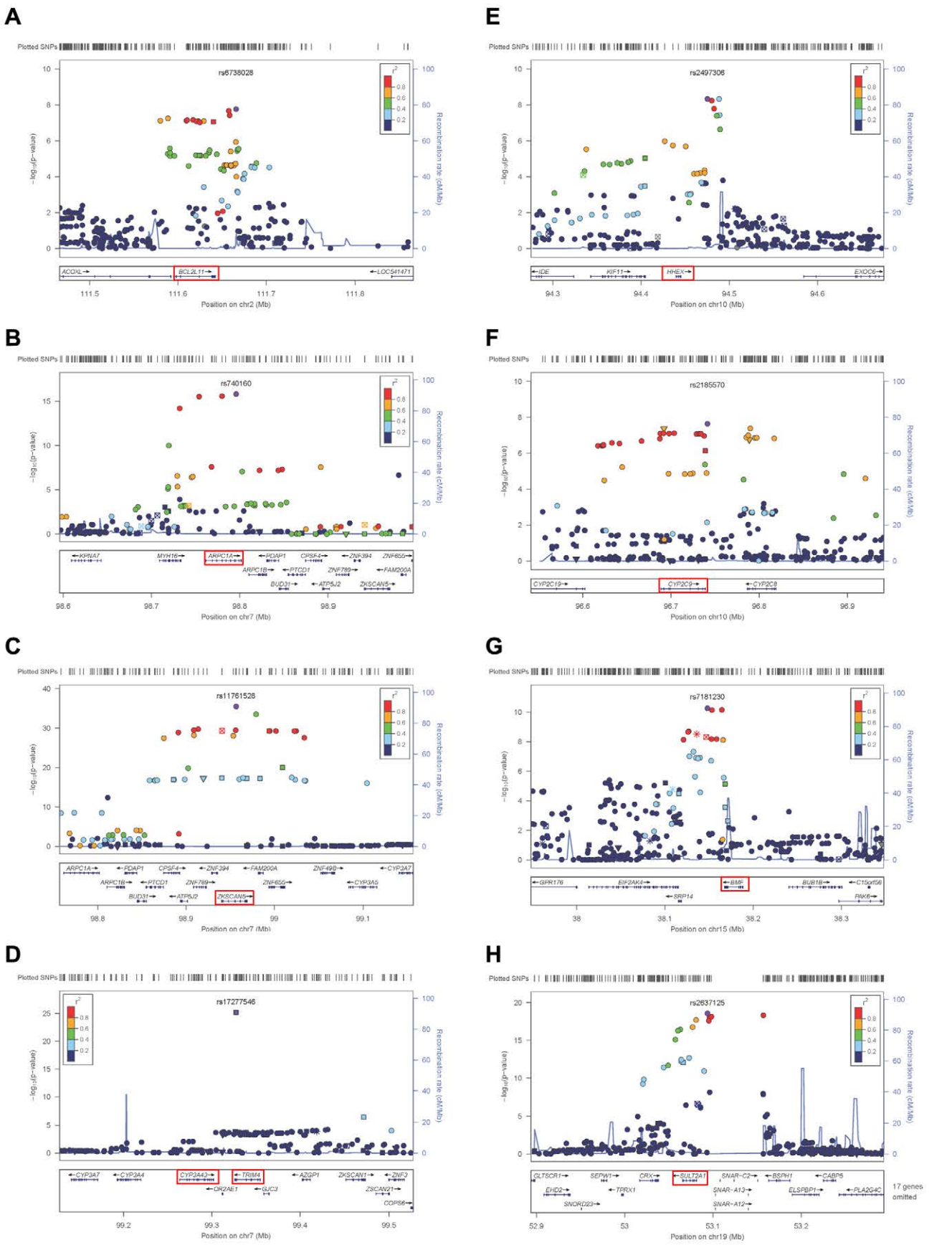

A
rs6738028
ACOXL
BCL2L11
LOC541471
Position on chr2 (Mb)
B
rs740160
KRNA7
MYH16
ARPC1A
PDAP1
CPSF4
ZNF394
ZNF655
ARPC1B
PTCD1
ZNF789
FAM200A
BUD31
ATP5J2
ZKSCAN5
Position on chr7 (Mb)
C
rs11761528
ARPC1A
PDAP1
CPSF4
ZNF394
FAM200A
ZNF498
CYP3A7
ARPC1B
PTCD1
ZNF789
ZNF655
CYP3A5
BUD31
ATP5J2
ZKSCAN5
Position on chr7 (Mb)
D
rs17277546
CYP3A7
CYP3A4
CYP3A43
TRIM4
AZGP1
ZKSCAN1
ZNF3
OR2AE1
GJC3
ZSCAN21
COPS6
Position on chr7 (Mb)
E
rs2497306
IDE
KIF11
HHEX
EXOC6
Position on chr10 (Mb)
F
rs2185570
CYP2C19
CYP2C9
CYP2C8
Position on chr10 (Mb)
G
rs7181230
GPR176
IQCA1
EIF2AK4
BMF
BUB1B
C15orf56
SRP14
PAK6
Position on chr15 (Mb)
H
rs2637125
GLTSCR1
SEPW1
CRX
SULT2A1
SNAR-C2
BSPH1
CABP5
EHD2
TPRX1
SNAR-A13
ELSPBP1
PLA2G4C
SNORD23
SNAR-A12
17 genes omitted
Position on chr19 (Mb)
Plotted SNPs
Recombination rate (cM/Mb)

Figure 2. Regional linkage disequilibrium plots. For rs6738028 (A), rs740160 (B), rs11761528 (C), rs17277546 (D), rs2497306 (E), rs2185570 (F), rs7181230 (G), and rs2637125 (H). Note: p values from the conditional analysis were used for (a) and (b), both of them became genome-wide significant in the conditional analysis. Annotation key: ▲ - framestop or splice; ▼ - NonSynonymous; ■ - Synonymous or UTR; ●- nothing; *- TFBScons; ⊠ -MCS44 Placental.
doi:10.1371/journal.pgen.1002025.g002

gene expression levels in at least one tissue after accounting for multiple testing (Table 3). These specific transcript associations provide further evidence for the likely functional gene at each locus.

Further, we carried out gene ontology and pathway analyses using a gene set enrichment analysis (GSEA) approach in MAGENTA [30] which consists of four main steps: First, DNA variants, e.g. SNP, are mapped onto genes. Second, each gene is assigned a gene association score that is a function of its regional SNP association p-values. Third, confounding effects on gene association scores are identified and corrected for, without requiring genotype data. Fourth, a GSEA-like statistical test is applied to predefined biologically relevant gene sets to determine whether any of the gene sets are enriched for highly ranked gene association scores compared to randomly sampled gene sets of identical size from the genome. More details of these four steps are described in the method section. In this analysis, we identified three pathways which passed our significance threshold (false discovery rate (FDR)$<$0.05); xenobiotic metabolism with FDR = 0.001 (pathway database: KEGG and Ingenuity), retinoid X receptor (RXR) function with FDR = 0.003 (pathway database: Ingenuity), and linoleic acid metabolism with FDR = 0.02 (pathway database: KEGG) (Figure S1). The top significant genes with $p<5.0\times10^{-8}$ include *CYP3A4*, *CYP3A43*, *CYP3A5*, and *CYP3A7* on chromosome 7, and *CYP2C8* and *CYP2C9* on chromosome 10 for all three pathways, and *SULT2A1* for RXR pathway. The best index SNPs are rs17277546 for *CYP3A4* and *CYP3A43*, rs4646450 for *CYP3A5* and *CYP3A7*, rs2185570 for *CYP2C9*, rs11572169 for *CYP2C8*, and rs2637125 for *SULT2A1*. The full list of the genes in each of the three pathways and the best index SNPs for each gene are listed in Table S2. Three SNPs – rs17277546, rs2185570, and rs2637125 are the DHEAS-associated SNPs found in our meta-analysis. Both rs4646450 and rs11572169 were associated with DHEAS with p values of 8.8×10^{-17} and 4.8×10^{-8}, respectively, but become non-significant in the conditional meta-analysis because rs4646450 is in linkage disequilibrium (LD, $r^2 = 0.429$) with rs11761528 which is the most significant DHEAS-associated SNP while rs11572169 is in high LD ($r^2 = 0.778$) with rs2185570. Intriguingly, two pathways - xenobiotic metabolism and linoleic acid metabolism, have been linked to ageing in model organisms [31–36].

Discussion

This is the first meta-analysis of GWA studies on serum DHEAS in 14,846 Caucasian subjects. We found 8 common SNPs that implicate nearby genes that are independently associated with serum DHEAS concentrations and provide clues to its role in ageing.

Among the genes identified, *SULT2A1*, a specialized sulphotransferase which converts DHEA to DHEAS in the adrenal cortex, is an obvious candidate gene [3]. *SULT2A1* has a broad substrate specificity, which includes conversion of pregnenolone, 17α-hydroxypregnenolone, and DHEA to their respective sulphated products [37]. Once sulphated by *SULT2A1*, pregnenolone and 17α-hydroxypregnenolone are no longer available as substrates for *HSD3B2*. Therefore, *SULT2A1* sulphation of pregnenolone and 17α-hydroxypregnenolone removes these substrates from the mineralocorticoid and glucocorticoid biosynthetic pathways. This suggests that high levels of *SULT2A1* would ensure the formation of DHEAS [3].

Variation in *SULT2A1* expression has previously been associated with variation of DHEAS concentration [38]. The *SULT2A1* gene is predominantly expressed in the adrenal cortex and to a lesser extent in the liver. We found that rs2547231 ($p = 1.76\times10^{-17}$), located 12 kb downstream of *SULT2A1*, was strongly associated with expression levels of *SULT2A1* in human liver tissues. Although this SNP is not the most strongly associated with serum DHEAS, it is itself in strong LD with the most significant SNP rs2637125 ($r^2 = 0.658$). However, we did not find a significant association with *SULT2A1* expression levels in LCL, skin, and adipose tissues, suggesting a tissue specific effect. The *SULT2B1b* is also reported to play a role in sulphation of DHEA, but in comparison, the strongest signal from that genomic region was rs10417472 with a p = 0.06. In contrast, enzymatic removal of the sulphate group from DHEAS to form DHEA is performed by steroid sulphatase gene (*STS*), but that gene is on the X chromosome and so was not assessed in this meta-analysis.

CYP2C9 is an important cytochrome P450 enzyme, accounts for approximately 17–20% of the total P450 content in human liver, and catalyzes many reactions involved in drug metabolism as well as synthesis of cholesterol, steroids and other lipids [39]. We found that rs2185570 located in the *CYP2C9* gene region is associated

Table 3. Association between DHEAS-associated SNPs and related gene expression levels in different human tissues.

Gene	Chr	SNP (effect allele)	Position	LCL* (n = 777)		Adipose tissue* (n = 776)		Skin tissue* (n = 667)		Liver tissue¶ (n = 427)
				Beta (SE)	P value	Beta (SE)	P value	Beta (SE)	P value	P value
BCL2L11	2	rs6738028 (G)	111665798	0.07 (0.02)	0.0003	0.02 (0.005)	0.001	−0.00004 (0.005)	0.99	Not available
TRIM4	7	rs17277546 (A)	99327507	0.15 (0.04)	0.0001	0.13(0.04)	0.002	0.10(0.04)	0.01	Not available
SULT2A1	19	rs2637125 (A)/ rs2547231**	53093705	0.0006 (0.007)	0.93	−0.009(0.007)	0.19	0.02(0.007)	0.01	2.16×10^{-11}

*from MuTHER consortium and beta (SE) were from linear regression modelling; LCL – lymphoblastoid cell lines.
¶from reference 27 and effect size was not reported.
**p value in liver expression is for rs2547231, data is not available for rs2637125, but two SNPs are in strong LD ($r^2 = 0.658$).
doi:10.1371/journal.pgen.1002025.t003

with serum DHEAS concentrations. This SNP is in strong LD with rs4086116 and rs4917639 ($r^2 = 0.67$ for both) which were found to be associated with acenocoumarol [40] and warfarin maintenance dosage [41] respectively in recent GWAS.

Two other cytochrome P450 enzymes – *CYP11A1* and *CYP17A1*, are two important enzymes which are required in the synthesis of DHEAS in the adrenal gland [3], however, the strongest signals in the genomic region were rs2930306 with p = 0.29 for *CYP11A1* and rs4919686 with p = 0.04 for *CYP17A1*.

The decline in serum DHEAS concentrations with increasing age has been proposed as a putative biomarker of ageing [21]. We found that two putative ageing genes – *BCL2L11* and *BMF* [42] are associated with serum DHEAS concentrations. Both of them encode proteins which belong to the *BCL2* family and act as anti- or pro-apoptotic regulators that are involved in a wide variety of cellular activities. *BCL2L11* has been implicated in chronic lymphocytic leukaemia (rs17483466, $P = 2.36 \times 10^{-10}$) [43] and follicular lymphoma (rs3789068, P for trend = 0.0004) [44]. The DHEAS-associated SNP rs6738028 is not however one of the same SNPs associated with lymphocytic leukaemia and follicular lymphoma nor is it in LD with them. Nevertheless, rs6738028 is strongly associated with *BCL2L11* gene expression levels in both LCL and adipose tissues, suggesting its putative regulatory role. There is relatively little data on the human *BMF* gene or the protein product, but $Bmf^{-/-}$ mice show altered immune and hematopoietic phenotypes as well as defects in uterovaginal development. However, we were not able to detect any association between rs7181230 and the expression levels of *BMF* in the tissues we studied.

HHEX encodes a member of the homeobox family of transcription factors, many of which are involved in developmental processes. This gene has been found to be associated with type 2 diabetes by several recent GWAS [45–51]. The risk alleles of the diabetes-associated SNPs rs1111875 and rs5015480 are associated with low serum DHEAS concentrations although the p values (p = 0.0009 for both SNPs) didn't reach to the GWAS significance level. This is consistent with the observation in which the low serum DHEAS concentrations were associated with insulin resistance [6,7]. The identified DHEAS-associated SNP rs2497306 is in moderate LD with rs1111875 and rs5015480 ($r^2 = 0.38$). And the major allele of rs2497306 is associated with increasing serum DHEAS concentrations. The reason for the observed association is unknown. Studies showed that insulin infusion increases the metabolic clearance of DHEA and DHEAS [52,53], resulting in decreased DHEA and DHEAS concentrations, and DHEA administration significantly enhances insulin sensitivity attenuating the age-related decline in glucose tolerance [54], partly explaining why the diabetes-associated gene is also associated with DHEAS. Interestingly, *HHEX* null mice show cardiovascular, endocrine, liver, muscle, nervous system, and metabolic phenotypes, suggesting extensive multisystem roles for the protein product of this gene. The findings could help dissect causal pathways for the observed associations between DHEAS, insulin resistance, age-related decline in glucose tolerance [54], and other age related phenotypes [55].

Three identified DHEAS-associated SNPs on chromosome 7 (Figure S2), which were independent, and 161 kb downstream (rs740160) and 370 kb upstream (rs17277546) apart from rs11761528 which is located in the middle of the region, are located in four genes - *ZKSCAN5*, *ARPC1A*, *and TRIM4/CYP3A43*. *ZKSCAN5* encodes a zinc finger protein of the Kruppel family and is expressed ubiquitously in adult and fetal tissues with the strongest expression in testis [56]. rs11761528 is located in the intron of the *ZKSCAN5* gene. It is the strongest signal we found and explains 1% of the total variance of serum DHEAS concentration alone. *ARPC1A* encodes one of seven subunits of the human Arp2/3 protein complex which has been implicated in actin polymerization and filament assembly in cells [57]. *TRIM4* encodes a member of the tripartite motif (TRIM) family whereas *CYP3A43* is another cytochrome P450 enzyme. The potential mechanisms for the association are unknown, but we found that rs17277546 is strongly associated with expression levels of *TRIM4* not *CYP3A43*, suggesting *TRIM4* is the possible candidate for DHEAS. However, rs17277546 is the best index SNP for both *CYP3A43* and *CYP3A4* genes in the pathway analysis, indicating a fine mapping in this region is needed to reveal the potential mechanism for the association. Further, the region harbours many other genes including *CYP3A7* which has been reported to increase the clearance of DHEA and DHEAS [58] and a common haplotype polymorphism in the gene has been associated with DHEAS [59,60]. However, none of the DHEAS-associated SNPs are associated with its expression levels in the tissues we studied, and the best index SNP rs4646450 for *CYP3A7* found in our pathway analysis is in LD with rs11761528 and become non-significant in the conditional analysis.

In the pathway analysis, two DHEAS-associated SNPs (rs2185570 and rs17277546) were contained in all three pathways we found and one SNP (rs2637125) was contained in the RXR function pathway. Intriguingly, components of the xenobiotic metabolism pathway have been linked to ageing in model organisms, for example, age-associated changes in expression of genes involved in xenobiotic metabolism have been identified in rats [31,32], up-regulation of xenobiotic detoxification genes has been observed in long-lived mutant mice [33], and adrenal xenobiotic-metabolizing activities increase with ageing in guinea pigs [34]. Furthermore, linoleic acid metabolism has also been linked to changes with ageing in rat cardiac muscle [35] and in human skin fibroblasts [36]. Taken together, these findings suggest that molecular pathways involved in ageing and longevity may also underlie DHEAS regulation, suggesting shared genetic components in both processes and corroborating a role for DHEAS as a marker of biological ageing.

In summary, this first GWAS identified eight independent SNPs associated with serum DHEAS concentrations. The related genes have various associations with steroid hormone metabolism, comorbidities of ageing including type 2 diabetes, lymphoma, actin filament assembly, drug and xenobiotic metabolism, and zinc fingers - suggesting a wider functional role for DHEAS than previously thought.

Methods

Study population

Seven study samples contributed to this meta-analysis of GWA studies on serum DHEAS concentrations, comprising a total of 14,846 men and women of Caucasian origin. The consortium was made up of populations from TwinsUK (n = 4,906), Framingham Heart Study (FHS) (n = 3,183), SHIP (n = 1,832), Rotterdam Study (RS1) (n = 1,597), InCHIANTI (n = 1,182), Health ABC (n = 1,222), and GOOD (n = 924). Full details can be found in Text S1.

DHEAS methods

Blood samples were collected from each of the study participants either after overnight fasting or non-fasting and the serum levels of DHEAS were measured by either immunoassay or liquid chromatography tandem mass spectrometry (LC-MS/MS) methods (Text S1). Because the distribution of the serum DHEAS

levels was skewed, we log transformed the concentrations and the transformed data used in the subsequent analysis.

Genotyping and imputation

Seven study populations were genotyped using a variety of genotyping platforms including Illumina (HumanHap 317k, 550k, 610k, 1M-Duo BeadChip) and Affymetrix (array 500K, 6.0). Each cohort followed a strict quality control on the genotyping data. More details on the quality control and filtering criteria can be found in Text S1. In order to increase genomic coverage and allow the evaluation of the same SNPs across as many study populations as possible, each study imputed genotype data based on the HapMap CEU Build 36. Algorithms were used to infer unobserved genotypes in a probabilistic manner in either MACH (http://www.sph.umich.edu/csg/abecasis/MACH), or IMPUTE [61]. We exclude non-genotyped SNPs with an imputation quality score <0.2 and SNPs with allele frequency <0.01 from meta-analysis.

Statistical method

Each study performed genome-wide association testing for serum concentrations of DHEAS across approximately 2.5 million SNPs under an additive genetic model separately in men and women (Text S1). The analyses were adjusted for age. In addition, the association testing was performed in the combined men and women data with adjustment for age and sex. Studies used PLINK, GenABEL, SNPTEST, QUICKTEST, or MERLIN for the association testing. The summary results from each cohort were meta-analyzed by Z-score pooling method implemented in Metal (http://www.sph.umich.edu/csg/abecasis/metal/). We chose this method to minimize the impact of the different assays used for serum DHEAS measurements. Specifically, for each study, we converted the two-sided P value after adjustment for population stratification by the genomic control method to a Z statistic that was signed to reflect the direction of the association given the reference allele. Each Z score was then weighted; the squared weights were chosen to sum to 1, and each sample-specific weight was proportional to the square root of the effective number of individuals in the sample. We summed the weighted Z statistics across studies and converted the summary Z score to a two-sided P value. We also used I^2 index to assess between-study heterogeneity and the inverse variance weighted method to estimate the effect size. Genome-wide significance was defined as $p<5\times10^{-8}$. The association between the DHEAS-associated SNPs and the related gene expression levels in MuTHER data were examined by mixed linear regression modelling which takes both family structure and batch effects into account. The significance was defined as $p<0.006$ after accounting for multiple testing (Bonferroni method, correcting 9 independent tests).

Pathway analysis. Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) was used to explore pathway-based associations in the full GWAS dataset. MAGENTA implements a gene set enrichment analysis (GSEA) based approach, the methodology of which is described in Segrè et al [30]. Briefly, each gene in the genome is mapped to a single index SNP with the lowest P-value within a 110 kb upstream, 40 kb downstream window. This P-value, representing a gene score, is then corrected for confounding factors such as gene size, SNP density and LD-related properties in a regression model. Genes within the HLA-region were excluded from analysis due to difficulties in accounting for gene density and LD patterns. Each mapped gene in the genome is then ranked by its adjusted gene score. At a given significance threshold (95th and 75th percentiles of all gene scores), the observed number of gene scores in a given pathway, with a ranked score above the specified threshold percentile, is calculated. This observed statistic is then compared to 1,000,000 randomly permuted pathways of identical size. This generates an empirical GSEA P-value for each pathway. Significance was determined when an individual pathway reached a false discovery rate (FDR)<0.05 in either analysis. In total, 2529 pathways from Gene Ontology, PANTHER, KEGG and Ingenuity were tested for enrichment of multiple modest associations with serum DHEAS levels.

Ethics statement

All studies were approved by local ethics committees and all participants provided written informed consent as stated in Text S1.

Supporting Information

Figure S1 Three pathways which were associated with DHEAS. The genes which are near the DHEAS-associated SNPs are highlighted by red circles. a. Xenobiotic metabolism pathway; b. Retinoid X receptor (RXR) function pathway; c. Linoleic acid metabolism pathway; d. Legends for the pathway figures. The pathway figures were made using MetaCore from GeneGo (http://www.genego.com/metacore.php).
(TIF)

Figure S2 Regional linkage disequilibrium plots for three SNPs on chromosome 7 in one plot.
(TIF)

Table S1 87 SNPs associated with DHEAS in men, women, and combined meta-analysis with $p<5\times10^{-8}$.
(XLS)

Table S2 Pathway analysis results – list of all pathways, significant pathways, and significant genes with the best index SNPs.
(XLS)

Text S1 Supplementary Note.
(DOC)

Acknowledgments

TwinsUK: The study was funded by the Wellcome Trust, Arthritis Research Campaign, European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F2-2008-201865-GEFOS and Seventh Framework Programme grant 200800 Treat OA/(FP7/2007-2013), ENGAGE project grant agreement HEALTH-F4-2007-201413, and the FP-5 GenomEUtwin Project (QLG2-CT-2002-01254). The study also receives support from the UK Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. TDS is an NIHR senior investigator. The project also received support from a Biotechnology and Biological Sciences Research Council project grant (G20234). The authors acknowledge the funding and support of the National Eye Institute (NEI) via an NIH/Center for Inherited Disease Research (CIDR) genotyping project. We thank the staff from the Genotyping Facilities at the Wellcome Trust Sanger Institute for sample preparation, quality control, and genotyping; Le Centre National de Génotypage, France, for genotyping; Duke University, NC, USA, for genotyping; and the Finnish Institute of Molecular Medicine, Finnish Genome Center, University of Helsinki. Genotyping was also done by CIDR as part of an NEI/NIH project grant.

SHIP: Computing resources have been made available by the Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities (HLRB project h1231). SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural

Affairs as well as the Social Ministry of the Federal State of Mecklenburg - West Pomerania. Genome- wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg West Pomerania. This work is also part of the research project Greifswald Approach to Individualized Medicine (GANI_MED). The GANI_MED consortium is funded by the Federal Ministry of Education and Research and the Ministry of Cultural Affairs of the Federal State of Mecklenburg – West Pomerania (03IS2061A). The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG. The DHEAS reagents used were sponsored by Siemens Healthcare Diagnostics, Eschborn, formerly DPC Biermann GmbH, Bad Nauheim, Germany. Novo Nordisc provided partial grant support for the determination of serum samples and data analysis.

Health ABC: This study was supported by National Institute on Aging contracts N01-AG-6-2101, N01-AG-6-2103, and N01-AG-6-2106. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C." This research was supported (in part) by the Intramural Research Program of the NIH, National Institute on Aging.

RS1: The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI) - Netherlands Consortium of Healthy Aging (NCHA) project nr. 050-060-810, and funding from the European Commision (HEALTH-F2-2008-201865, GEFOS; HEALTH-F2-2008-35627, TREAT-OA). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. We thank Pascal Arp, Mila Jhamai, Dr Michael Moorhouse, Marijn Verkerk, and Sander Bervoets for their help in creating the GWAS database. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practioners and pharmacists. We would like to thank Dr. Tobias A. Knoch, Karol Estrada, Luc V. de Zeeuw, Anis Abuseiris, and Rob de Graaf as well as their institutions the Erasmus Computing Grid, Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technology under grants #01 AK 803 A-H and # 01 IG 07015 G for access to their grid resources.

FHS: DHEAS-genotype genome-wide association analyses were supported by the National Institute of Aging (Genetics of Reproductive Life Period and Health Outcomes, R21AG032598; JMM, KL, DPK, WVZ). "The Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study Contract No. N01-HC-25195 and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. A portion of this research was conducted using the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center." DPK received support from NIH (Grant R01 AR/AG 41398).

GOOD: Financial support was received from the Swedish Research Council (K2010-54X-09894-19-3, 2006-3832), the Swedish Foundation for Strategic Research, the ALF/LUA research grant in Gothenburg, the Lundberg Foundation, the Torsten and Ragnar Söderberg's Foundation, Petrus and Augusta Hedlunds Foundation, the Västra Götaland Foundation, the Göteborg Medical Society, the Novo Nordisk foundation and the European Commission grant HEALTH-F2-2008-201865-GEFOS. We would like to thank Dr. Tobias A. Knoch, Luc V. de Zeeuw, Anis Abuseiris, and Rob de Graaf as well as their institutions the Erasmus Computing Grid, Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technology under grants #01 AK 803 A-H and # 01 IG 07015 G for access to their grid resources. We would also like to thank Karol Estrada, Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands for advice regarding the grid resources.

InCHIANTI: The InCHIANTI study baseline (1998–2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336); the InCHIANTI Follow-up 1 (2001–2003) was funded by the U.S. National Institute on Aging (Contracts: N.1-AG-1-1 and N.1-AG-1-2111); the InCHIANTI Follow-ups 2 and 3 studies (2004–2010) were financed by the U.S. National Institute on Aging (Contract: N01-AG-5-0002); supported in part by the Intramural research program of the National Institute on Aging, National Institutes of Health, Baltimore, Maryland. JRBP is funded by a Sir Henry Wellcome Postdoctoral Fellowship (092447/Z/10/Z).

Author Contributions

Study design: TD Spector, G Zhai, M Mangino, A Hofman, AG Uitterlinden, AD Coviello, C Ohlsson, A Teumer, L Ferrucci, M Maggio, TM Frayling, H Wallaschofski, JM Murabito, A Koster, TB Harris, DP Kiel. Sample collection and preparation: TD Spector, D Glass, A Hofman, FH de Jong, C Ohlsson, M Lorentzon, A Eriksson, L Vandenput, HK Kroemer, H Völzke, H Wallaschofski, S Bhasin, L Ferrucci, M Marcello, F Rivadeneira. Statistical analyses: G Zhai, M Mangino, N Soranzo, KS Small, ÅK Hedman, E Grundberg, A Nica, L Stolk, L Vandenput, J Eriksson, A Teumer, F Ernst, JRB Perry, KL Lunetta, WV Zhuang, Y Liu, F Rivadeneira. Manuscript writing: G Zhai, TD Spector, M Mangino, L Stolk, F de Jong, AD Coviello, C Ohlsson, JM Murabito, A Koster, JRB Perry. Review and revision of the manuscript: G Zhai, M Mangino, TD Spector, N Soranzo, JT Bell, KS Small, A Hofman, E Grundberg, A Nica, SG Wilson, L Stolk, A Teumer, H Koster, H Völzke, F Ernst, R Haring, HK Kroemer, H Wallaschofski, FH de Jong, AD Coviello, C Ohlsson, L Vandenput, M Lorentzon, J Eriksson, A Eriksson, JRB Perry, D Melzer, TM Frayling, M Maggio, L Ferrucci, A Murray, JM Murabito, ÅK Hedman, RS Vasan, DP Kiel, A Koster, TB Harris, Y Liu, BWJH Pennix, I Miljkovic, KL Lunetta, L Xi. Intellectual input: SG Wilson, JT Bell, C Ohlsson, R Haring, A Teumer, JRB Perry, TM Frayling, A Murray, H Wallaschofski.

References

1. Labrie F (1991) Intracrinology. Mol Cell Endocrinol 78: C113–118.
2. Thomas G, Frenoy N, Legrain S, Sebag-Lanoe R, Baulieu EE, Debuire B (1994) Serum dehydroepiandrosterone sulfate levels as an individual marker. J Clin Endocrinol Metab 79: 1273–1276.
3. Rainey WE, Carr BR, Sasano H, Suzuki T, Mason JI (2002) Dissecting human adrenal androgen production. Trends Endocrinol Metab 13: 234–239.
4. Leowattana W (2001) DHEA(S): the fountain of youth. J Med Assoc Thai 84 Suppl 2: S605–612.
5. Celec P, Starka L (2003) Dehydroepiandrosterone - is the fountain of youth drying out? Physiol Res 52: 397–407.
6. Mottl R, Cerman J (2004) A relationship between dehydroepiandrosterone sulphate and insulin resistance in obese men and women. Vnitr Lek 50: 923–929.
7. Brennan K, Huang A, Azziz R (2009) Dehydroepiandrosterone sulfate and insulin resistance in patients with polycystic ovary syndrome. Fertil Steril 91: 1848–1852.
8. Suzuki M, Kanazawa A, Hasegawa M, Hattori Y, Harano Y (1999) A close association between insulin resistance and dehydroepiandrosterone sulfate in subjects with essential hypertension. Endocr J 46: 521–528.
9. Haring R, Volzke H, Felix SB, Schipf S, Dorr M, Rosskopf D, Nauck M, Schofl C, Wallaschofski H (2009) Prediction of metabolic syndrome by low serum testosterone levels in men: results from the study of health in Pomerania. Diabetes 58: 2027–2031.
10. Szathmari M, Szucs J, Feher T, Hollo I (1994) Dehydroepiandrosterone sulphate and bone mineral density. Osteoporos Int 4: 84–88.
11. Miklos S (1995) Dehydroepiandrosterone sulphate in the diagnosis of osteoporosis. Acta Biomed Ateneo Parmense 66: 139–146.
12. Sun AJ, Jing T, Heymsfield SB, Phillips GB (2003) Relationship of leptin and sex hormones to bone mineral density in men. Acta Diabetol 40 Suppl 1: S101–105.
13. Sanders JL, Cappola AR, Arnold AM, Boudreau RM, Chaves PH, Robbins J, Cushman M, Newman AB (2010) Concurrent Change in Dehydroepiandros-

terone Sulfate and Functional Performance in the Oldest Old: Results From the Cardiovascular Health Study All Stars Study. J Gerontol A Biol Sci Med Sci.
14. Barrett-Connor E, Khaw KT, Yen SS (1986) A prospective study of dehydroepiandrosterone sulfate, mortality, and cardiovascular disease. N Engl J Med 315: 1519–1524.
15. Jansson JH, Nilsson TK, Johnson O (1998) von Willebrand factor, tissue plasminogen activator, and dehydroepiandrosterone sulphate predict cardiovascular death in a 10 year follow up of survivors of acute myocardial infarction. Heart 80: 334–337.
16. Mazat L, Lafont S, Berr C, Debuire B, Tessier JF, Dartigues JF, Baulieu EE (2001) Prospective measurements of dehydroepiandrosterone sulfate in a cohort of elderly subjects: relationship to gender, subjective health, smoking habits, and 10-year mortality. Proc Natl Acad Sci U S A 98: 8145–8150.
17. Cappola AR, O'Meara ES, Guo W, Bartz TM, Fried LP, Newman AB (2009) Trajectories of dehydroepiandrosterone sulfate predict mortality in older adults: the cardiovascular health study. J Gerontol A Biol Sci Med Sci 64: 1268–1274.
18. Ohlsson C, Labrie F, Barrett-Connor E, Karlsson MK, Ljunggren O, Vandenput L, Mellstrom D, Tivesten A (2010) Low serum levels of dehydroepiandrosterone sulfate predict all-cause and cardiovascular mortality in elderly Swedish men. J Clin Endocrinol Metab 95: 4406–4414.
19. Barrett-Connor E, Goodman-Gruen D (1995) The epidemiology of DHEAS and cardiovascular disease. Ann N Y Acad Sci 774: 259–270.
20. Phillips AC, Carroll D, Gale CR, Lord JM, Arlt W, Batty GD (2010) Cortisol, DHEAS, their ratio and the metabolic syndrome: evidence from the Vietnam Experience Study. Eur J Endocrinol 162: 919–923.
21. Leowattana W (2004) DHEAS as a new diagnostic tool. Clin Chim Acta 341: 1–15.
22. Nestler JE, Whitfield JB, Williams TY, Zhu G, Condon J, Kirk KM, Heath AC, Montgomery GW, Martin NG (2002) Genetics of serum dehydroepiandrosterone sulfate and its relationship to insulin in a population-based cohort of twin subjects. J Clin Endocrinol Metab 87: 682–686.
23. Meikle AW, Stringham JD, Woodward MG, Bishop DT (1988) Heritability of variation of plasma cortisol levels. Metabolism 37: 514–517.
24. An P, Rice T, Gagnon J, Hong Y, Leon AS, Skinner JS, Wilmore JH, Bouchard C, Rao DC (2001) Race differences in the pattern of familial aggregation for dehydroepiandrosterone sulfate and its responsiveness to training in the HERITAGE Family Study. Metabolism 50: 916–920.
25. Rotter JI, Wong FL, Lifrak ET, Parker LN (1985) A genetic component to the variation of dehydroepiandrosterone sulfate. Metabolism 34: 731–736.
26. Davey Smith G, Ebrahim S (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 32: 1–22.
27. Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997–1004.
28. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, et al. (2011) The Architecture of Gene Regulatory Variation across Multiple Human Tissues: The MuTHER Study. PLoS Genet 7: e1002003. doi:10.1371/journal.pgen.1002003.
29. Schadt EE, Molony C, Chudin E, et al. (2008) Mapping the genetic architecture of gene expression in human liver. PLoS Biol 6: e107. doi:10.1371/journal.pbio.0060107.
30. Segre AV, Groop L, Mootha VK, Daly MJ, Altshuler D (2010) Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. PLoS Genet 6: e1001058. doi:10.1371/journal.pgen.1001058.
31. Mori K, Blackshear PE, Lobenhofer EK, et al. (2007) Hepatic transcript levels for genes coding for enzymes associated with xenobiotic metabolism are altered with age. Toxicol Pathol 35: 242–251.
32. Lee JH, Banerjee A, Ueno Y, Ramaiah SK (2008) Potential relationship between hepatobiliary osteopontin and peroxisome proliferator-activated receptor alpha expression following ethanol-associated hepatic injury in vivo and in vitro. Toxicol Sci 106: 290–299.
33. Amador-Noguez D, Dean A, Huang W, Setchell K, Moore D, Darlington G (2007) Alterations in xenobiotic metabolism in the long-lived Little mice. Aging Cell 6: 453–470.
34. Yuan BB, Tchao R, Voigt JM, Colby HD (2001) Maturational changes in CYP2D16 expression and xenobiotic metabolism in adrenal glands from male and female guinea pigs. Drug Metab Dispos 29: 194–199.
35. Lopez Jimenez JA, Bordoni A, Lorenzini A, Rossi CA, Biagi PL, Hrelia S (1997) Linoleic acid metabolism in primary cultures of adult rat cardiomyocytes is impaired by aging. Biochem Biophys Res Commun 237: 142–145.
36. Raederstorff D, Loechleiter V, Moser U (1995) Polyunsaturated fatty acid metabolism of human skin fibroblasts during cellular aging. Int J Vitam Nutr Res 65: 51–55.
37. Weinshilboum RM, Otterness DM, Aksoy IA, Wood TC, Her C, Raftogianis RB (1997) Sulfation and sulfotransferases 1: Sulfotransferase molecular biology: cDNAs and genes. Faseb J 11: 3–14.
38. Suzuki T, Sasano H, Takeyama J, Kaneko C, Freije WA, Carr BR, Rainey WE (2000) Developmental changes in steroidogenic enzymes in human postnatal adrenal cortex: immunohistochemical studies. Clin Endocrinol (Oxf) 53: 739–747.
39. Omiecinski CJ, Remmel RP, Hosagrahara VP (1999) Concise review of the cytochrome P450s and their roles in toxicology. Toxicol Sci 48: 151–156.
40. Teichert M, Eijgelsheim M, Rivadeneira F, Uitterlinden AG, van Schaik RH, Hofman A, De Smet PA, van Gelder T, Visser LE, Stricker BH (2009) A genome-wide association study of acenocoumarol maintenance dosage. Human molecular genetics 18: 3758–3768.
41. Takeuchi F, McGinnis R, Bourgeois S, et al. (2009) A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. PLoS Genet 5: e1000433. doi:10.1371/journal.pgen.1000433.
42. Lukiw WJ, Bazan NG (2010) Inflammatory, apoptotic, and survival gene signaling in Alzheimer's disease. A review on the bioactivity of neuroprotectin D1 and apoptosis. Mol Neurobiol 42: 10–16.
43. Di Bernardo MC, Crowther-Swanepoel D, Broderick P, et al. (2008) A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. Nat Genet 40: 1204–1210.
44. Morton LM, Purdue MP, Zheng T, et al. (2009) Risk of non-Hodgkin lymphoma associated with germline variation in genes that regulate the cell cycle, apoptosis, and lymphocyte development. Cancer Epidemiol Biomarkers Prev 18: 1259–1270.
45. Voight BF, Scott LJ, Steinthorsdottir V, et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42: 579–589.
46. Takeuchi F, Serizawa M, Yamamoto K, et al. (2009) Confirmation of multiple risk Loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. Diabetes 58: 1690–1699.
47. Zeggini E, Scott LJ, Saxena R, et al. (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 40: 638–645.
48. Saxena R, Voight BF, Lyssenko V, et al. (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316: 1331–1336.
49. Scott LJ, Mohlke KL, Bonnycastle LL, et al. (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316: 1341–1345.
50. Zeggini E, Weedon MN, Lindgren CM, et al. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316: 1336–1341.
51. Sladek R, Rocheleau G, Rung J, et al. (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445: 881–885.
52. Lavallee B, Provost PR, Kahwash Z, Nestler JE, Belanger A (1997) Effect of insulin on serum levels of dehydroepiandrosterone metabolites in men. Clin Endocrinol (Oxf) 46: 93–100.
53. Nestler JE, Beer NA, Jakubowicz DJ, Beer RM (1994) Effects of a reduction in circulating insulin by metformin on serum dehydroepiandrosterone sulfate in nondiabetic men. J Clin Endocrinol Metab 78: 549–554.
54. Bates GW, Jr., Egerman RS, Umstot ES, Buster JE, Casson PR (1995) Dehydroepiandrosterone attenuates study-induced declines in insulin sensitivity in postmenopausal women. Ann N Y Acad Sci 774: 291–293.
55. Bogue CW, Zhang PX, McGrath J, Jacobs HC, Fuleihan RL (2003) Impaired B cell development and function in mice with a targeted disruption of the homeobox gene Hex. Proc Natl Acad Sci U S A 100: 556–561.
56. Dreyer SD, Zheng Q, Zabel B, Winterpacht A, Lee B (1999) Isolation, characterization, and mapping of a zinc finger gene, ZFP95, containing both a SCAN box and an alternatively spliced KRAB A domain. Genomics 62: 119–122.
57. Welch MD, DePace AH, Verma S, Iwamatsu A, Mitchison TJ (1997) The human Arp2/3 complex is composed of evolutionarily conserved subunits and is localized to cellular regions of dynamic actin filament assembly. J Cell Biol 138: 375–384.
58. Miller KK, Cai J, Ripp SL, Pierce WM, Jr., Rushmore TH, Prough RA (2004) Stereo- and regioselectivity account for the diversity of dehydroepiandrosterone (DHEA) metabolites produced by liver microsomal cytochromes P450. Drug Metab Dispos 32: 305–313.
59. Smit P, van Schaik RH, van der Werf M, van den Beld AW, Koper JW, Lindemans J, Pols HA, Brinkmann AO, de Jong FH, Lamberts SW (2005) A common polymorphism in the CYP3A7 gene is associated with a nearly 50% reduction in serum dehydroepiandrosterone sulfate levels. J Clin Endocrinol Metab 90: 5313–5316.
60. Goodarzi MO, Xu N, Azziz R (2008) Association of CYP3A7*1C and serum dehydroepiandrosterone sulfate levels in women with polycystic ovary syndrome. J Clin Endocrinol Metab 93: 2909–2912.
61. Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 39: 906–913.